UNITED STATES AIR FORCE

SUMMER RESEARCH PROGRAM -- 1996

GRADUATE STUDENT RESEARCH PROGRAM FINAL REPORTS

VOLUME 7A

ARMSTRONG LABORATORY

RESEARCH & DEVELOPMENT LABORATORIES

5800 Uplander Way

Culver City, CA 90230-6608

Program Director, RDL
Gary Moore

Program Manager, AFOSR
Major Linda Steel-Goodwin

Program Manager, RDL
Scott Licoscos

Program Administrator, RDL
Johnetta Thompson

Program Administrator, RDL
Rebecca Kelly

Submitted to:

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

Bolling Air Force Base

Washington, D.C.

December 1996

REPORT DOCUMENTATION PAGE

AFRL-SR-BL-TR-00-
0736

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructior ...] and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, includin ... for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
December, 1996

3. R

4. TITLE AND SUBTITLE
1996 Summer Research Program (SRP), Graduate Student Research Program (GSRP), Final Reports, Volume 7A, Armstrong Laboratory

5. FUNDING NUMBERS
F49620-93-C-0063

6. AUTHOR(S)
Gary Moore

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Research & Development Laboratories (RDL)
5800 Uplander Way
Culver City, CA 90230-6608

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Office of Scientific Research (AFOSR)
801 N. Randolph St.
Arlington, VA 22203-1977

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION AVAILABILITY STATEMENT
Approved for Public Release

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The United States Air Force Summer Research Program (USAF-SRP) is designed to introduce university, college, and technical institute faculty members, graduate students, and high school students to Air Force research. This is accomplished by the faculty members (Summer Faculty Research Program, (SFRP)), graduate students (Graduate Student Research Program (GSRP)), and high school students (High School Apprenticeship Program (HSAP)) being selected on a nationally advertised competitive basis during the summer intersession period to perform research at Air Force Research Laboratory (AFRL) Technical Directorates, Air Force Air Logistics Centers (ALC), and other AF Laboratories. This volume consists of a program overview, program management statistics, and the final technical reports from the GSRP participants at the Armstrong Laboratory.

14. SUBJECT TERMS
Air Force Research, Air Force, Engineering, Laboratories, Reports, Summer, Universities, Faculty, Graduate Student, High School Student

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to ***stay within the lines*** to meet ***optical scanning requirements.***

Block 1. Agency Use Only *(Leave blank)*.

Block 2. Report Date. Full publication date including day, month, and year, if available
(e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C - Contract
G - Grant
PE - Program Element
PR - Project
TA - Task
WU - Work Unit Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Name(s) and Address(es). Self-explanatory.

Block 10. Sponsoring/Monitoring Agency Report Number. *(If known)*

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with....; Trans. of....; To be published in.... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement. Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

DOD - See DoDD 5230.24, "Distribution Statements on Technical Documents."
DOE - See authorities.
NASA - See Handbook NHB 2200.2.
NTIS - Leave blank.

Block 12b. Distribution Code.

DOD - Leave blank.
DOE - Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.
Leave blank.
NASA - Leave blank.
NTIS -

Block 13. Abstract. Include a brief *(Maximum 200 words)* factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code *(NTIS only)*.

Blocks 17. - 19. Security Classifications. Self- explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.

PREFACE

Reports in this volume are numbered consecutively beginning with number 1. Each report is paginated with the report number followed by consecutive page numbers, e.g., 1-1, 1-2, 1-3; 2-1, 2-2, 2-3.

Due to its length, Volume 7 is bound in three parts, 7A, and 7B. Volume 7A contains #1-23. Volume 7B contains reports #24-35. The Table of Contents for Volume 7 is included in all parts.

This document is one of a set of 16 volumes describing the 1996 AFOSR Summer Research Program. The following volumes comprise the set:

VOLUME	TITLE
1	Program Management Report
	Summer Faculty Research Program (SFRP) Reports
2A & 2B	Armstrong Laboratory
3A & 3B	Phillips Laboratory
4	Rome Laboratory
5A , 5B & 5C	Wright Laboratory
6	Arnold Engineering Development Center, Wilford Hall Medical Center and Air Logistics Centers
	Graduate Student Research Program (GSRP) Reports
7A & 7B	Armstrong Laboratory
8	Phillips Laboratory
9	Rome Laboratory
10A & 10B	Wright Laboratory
11	Arnold Engineering Development Center, United States Air Force Academy, Wilford Hall Medical Center, and Wright Patterson Medical Center
	High School Apprenticeship Program (HSAP) Reports
12A & 12B	Armstrong Laboratory
13	Phillips Laboratory
14	Rome Laboratory
15A&15B	Wright Laboratory
16	Arnold Engineering Development Center

GSRP FINAL REPORT TABLE OF CONTENTS i-x

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

INTRODUCTION

The Summer Research Program (SRP), sponsored by the Air Force Office of Scientific Research (AFOSR), offers paid opportunities for university faculty, graduate students, and high school students to conduct research in U.S. Air Force research laboratories nationwide during the summer.

Introduced by AFOSR in 1978, this innovative program is based on the concept of teaming academic researchers with Air Force scientists in the same disciplines using laboratory facilities and equipment not often available at associates' institutions.

The Summer Faculty Research Program (SFRP) is open annually to approximately 150 faculty members with at least two years of teaching and/or research experience in accredited U.S. colleges, universities, or technical institutions. SFRP associates must be either U.S. citizens or permanent residents.

The Graduate Student Research Program (GSRP) is open annually to approximately 100 graduate students holding a bachelor's or a master's degree; GSRP associates must be U.S. citizens enrolled full time at an accredited institution.

The High School Apprentice Program (HSAP) annually selects about 125 high school students located within a twenty mile commuting distance of participating Air Force laboratories.

AFOSR also offers its research associates an opportunity, under the Summer Research Extension Program (SREP), to continue their AFOSR-sponsored research at their home institutions through the award of research grants. In 1994 the maximum amount of each grant was increased from $20,000 to $25,000, and the number of AFOSR-sponsored grants decreased from 75 to 60. A separate annual report is compiled on the SREP.

The numbers of projected summer research participants in each of the three categories and SREP "grants" are usually increased through direct sponsorship by participating laboratories.

AFOSR's SRP has well served its objectives of building critical links between Air Force research laboratories and the academic community, opening avenues of communications and forging new research relationships between Air Force and academic technical experts in areas of national interest, and strengthening the nation's efforts to sustain careers in science and engineering. The success of the SRP can be gauged from its growth from inception (see Table 1) and from the favorable responses the 1996 participants expressed in end-of-tour SRP evaluations (Appendix B).

AFOSR contracts for administration of the SRP by civilian contractors. The contract was first awarded to Research & Development Laboratories (RDL) in September 1990. After

completion of the 1990 contract, RDL (in 1993) won the recompetition for the basic year and four 1-year options.

2. PARTICIPATION IN THE SUMMER RESEARCH PROGRAM

The SRP began with faculty associates in 1979; graduate students were added in 1982 and high school students in 1986. The following table shows the number of associates in the program each year.

YEAR	SRP Participation, by Year			TOTAL
	SFRP	GSRP	HSAP	
1979	70			70
1980	87			87
1981	87			87
1982	91	17		108
1983	101	53		154
1984	152	84		236
1985	154	92		246
1986	158	100	42	300
1987	159	101	73	333
1988	153	107	101	361
1989	168	102	103	373
1990	165	121	132	418
1991	170	142	132	444
1992	185	121	159	464
1993	187	117	136	440
1994	192	117	133	442
1995	190	115	137	442
1996	188	109	138	435

Beginning in 1993, due to budget cuts, some of the laboratories weren't able to afford to fund as many associates as in previous years. Since then, the number of funded positions has remained fairly constant at a slightly lower level.

3. RECRUITING AND SELECTION

The SRP is conducted on a nationally advertised and competitive-selection basis. The advertising for faculty and graduate students consisted primarily of the mailing of 8,000 52-page SRP brochures to chairpersons of departments relevant to AFOSR research and to administrators of grants in accredited universities, colleges, and technical institutions. Historically Black Colleges and Universities (HBCUs) and Minority Institutions (MIs) were included. Brochures also went to all participating USAF laboratories, the previous year's participants, and numerous individual requesters (over 1000 annually).

RDL placed advertisements in the following publications: *Black Issues in Higher Education, Winds of Change,* and *IEEE Spectrum.* Because no participants list either *Physics Today* or *Chemical & Engineering News* as being their source of learning about the program for the past several years, advertisements in these magazines were dropped, and the funds were used to cover increases in brochure printing costs.

High school applicants can participate only in laboratories located no more than 20 miles from their residence. Tailored brochures on the HSAP were sent to the head counselors of 180 high schools in the vicinity of participating laboratories, with instructions for publicizing the program in their schools. High school students selected to serve at Wright Laboratory's Armament Directorate (Eglin Air Force Base, Florida) serve eleven weeks as opposed to the eight weeks normally worked by high school students at all other participating laboratories.

Each SFRP or GSRP applicant is given a first, second, and third choice of laboratory. High school students who have more than one laboratory or directorate near their homes are also given first, second, and third choices.

Laboratories make their selections and prioritize their nominees. AFOSR then determines the number to be funded at each laboratory and approves laboratories' selections.

Subsequently, laboratories use their own funds to sponsor additional candidates. Some selectees do not accept the appointment, so alternate candidates are chosen. This multi-step selection procedure results in some candidates being notified of their acceptance after scheduled deadlines. The total applicants and participants for 1996 are shown in this table.

1996 Applicants and Participants			
PARTICIPANT CATEGORY	TOTAL APPLICANTS	SELECTEES	DECLINING SELECTEES
SFRP	572	188	39
(HBCU/MI)	(119)	(27)	(5)
GSRP	235	109	7
(HBCU/MI)	(18)	(7)	(1)
HSAP	474	138	8
TOTAL	1281	435	54

4. SITE VISITS

During June and July of 1996, representatives of both AFOSR/NI and RDL visited each participating laboratory to provide briefings, answer questions, and resolve problems for both laboratory personnel and participants. The objective was to ensure that the SRP would be as constructive as possible for all participants. Both SRP participants and RDL representatives found these visits beneficial. At many of the laboratories, this was the only opportunity for all participants to meet at one time to share their experiences and exchange ideas.

5. HISTORICALLY BLACK COLLEGES AND UNIVERSITIES AND MINORITY INSTITUTIONS (HBCU/MIs)

Before 1993, an RDL program representative visited from seven to ten different HBCU/MIs annually to promote interest in the SRP among the faculty and graduate students. These efforts were marginally effective, yielding a doubling of HBCI/MI applicants. In an effort to achieve AFOSR's goal of 10% of all applicants and selectees being HBCU/MI qualified, the RDL team decided to try other avenues of approach to increase the number of qualified applicants. Through the combined efforts of the AFOSR Program Office at Bolling AFB and RDL, two very active minority groups were found, HACU (Hispanic American Colleges and Universities) and AISES (American Indian Science and Engineering Society). RDL is in communication with representatives of each of these organizations on a monthly basis to keep up with the their activities and special events. Both organizations have widely-distributed magazines/quarterlies in which RDL placed ads.

Since 1994 the number of both SFRP and GSRP HBCU/MI applicants and participants has increased ten-fold, from about two dozen SFRP applicants and a half dozen selectees to over 100 applicants and two dozen selectees, and a half-dozen GSRP applicants and two or three selectees to 18 applicants and 7 or 8 selectees. Since 1993, the SFRP had a two-fold applicant

increase and a two-fold selectee increase. Since 1993, the GSRP had a three-fold applicant increase and a three to four-fold increase in selectees.

In addition to RDL's special recruiting efforts, AFOSR attempts each year to obtain additional funding or use leftover funding from cancellations the past year to fund HBCU/MI associates. This year, 5 HBCU/MI SFRPs declined after they were selected (and there was no one qualified to replace them with). The following table records HBCU/MI participation in this program.

SRP HBCU/MI Participation, By Year				
YEAR	SFRP		GSRP	
	Applicants	Participants	Applicants	Participants
1985	76	23	15	11
1986	70	18	20	10
1987	82	32	32	10
1988	53	17	23	14
1989	39	15	13	4
1990	43	14	17	3
1991	42	13	8	5
1992	70	13	9	5
1993	60	13	6	2
1994	90	16	11	6
1995	90	21	20	8
1996	119	27	18	7

6. SRP FUNDING SOURCES

Funding sources for the 1996 SRP were the AFOSR-provided slots for the basic contract and laboratory funds. Funding sources by category for the 1996 SRP selected participants are shown here.

1996 SRP FUNDING CATEGORY	SFRP	GSRP	HSAP
AFOSR Basic Allocation Funds	141	85	123
USAF Laboratory Funds	37	19	15
HBCU/MI By AFOSR (Using Procured Addn'l Funds)	10	5	0
TOTAL	188	109	138

SFRP - 150 were selected, but nine canceled too late to be replaced.
GSRP - 90 were selected, but five canceled too late to be replaced (10 allocations for the ALCs were withheld by AFOSR.)
HSAP - 125 were selected, but two canceled too late to be replaced.

7. COMPENSATION FOR PARTICIPANTS

Compensation for SRP participants, per five-day work week, is shown in this table.

1996 SRP Associate Compensation

PARTICIPANT CATEGORY	1991	1992	1993	1994	1995	1996
Faculty Members	$690	$718	$740	$740	$740	$770
Graduate Student (Master's Degree)	$425	$442	$455	$455	$455	$470
Graduate Student (Bachelor's Degree)	$365	$380	$391	$391	$391	$400
High School Student (First Year)	$200	$200	$200	$200	$200	$200
High School Student (Subsequent Years)	$240	$240	$240	$240	$240	$240

The program also offered associates whose homes were more than 50 miles from the laboratory an expense allowance (seven days per week) of $50/day for faculty and $40/day for graduate students. Transportation to the laboratory at the beginning of their tour and back to their home destinations at the end was also reimbursed for these participants. Of the combined SFRP and

GSRP associates, 65 % (194 out of 297) claimed travel reimbursements at an average round-trip cost of $780.

Faculty members were encouraged to visit their laboratories before their summer tour began. All costs of these orientation visits were reimbursed. Forty-five percent (85 out of 188) of faculty associates took orientation trips at an average cost of $444. By contrast, in 1993, 58 % of SFRP associates took orientation visits at an average cost of $685; that was the highest percentage of associates opting to take an orientation trip since RDL has administered the SRP, and the highest average cost of an orientation trip. These 1993 numbers are included to show the fluctuation which can occur in these numbers for planning purposes.

Program participants submitted biweekly vouchers countersigned by their laboratory research focal point, and RDL issued paychecks so as to arrive in associates' hands two weeks later.

In 1996, RDL implemented direct deposit as a payment option for SFRP and GSRP associates. There were some growing pains. Of the 128 associates who opted for direct deposit, 17 did not check to ensure that their financial institutions could support direct deposit (and they couldn't), and eight associates never did provide RDL with their banks' ABA number (direct deposit bank routing number), so only 103 associates actually participated in the direct deposit program. The remaining associates received their stipend and expense payments via checks sent in the US mail.

HSAP program participants were considered actual RDL employees, and their respective state and federal income tax and Social Security were withheld from their paychecks. By the nature of their independent research, SFRP and GSRP program participants were considered to be consultants or independent contractors. As such, SFRP and GSRP associates were responsible for their own income taxes, Social Security, and insurance.

8. CONTENTS OF THE 1996 REPORT

The complete set of reports for the 1996 SRP includes this program management report (Volume 1) augmented by fifteen volumes of final research reports by the 1996 associates, as indicated below:

1996 SRP Final Report Volume Assignments

LABORATORY	SFRP	GSRP	HSAP
Armstrong	2	7	12
Phillips	3	8	13
Rome	4	9	14
Wright	5A, 5B	10	15
AEDC, ALCs, WHMC	6	11	16

APPENDIX A – PROGRAM STATISTICAL SUMMARY

A. Colleges/Universities Represented

Selected SFRP associates represented 169 different colleges, universities, and institutions, GSRP associates represented 95 different colleges, universities, and institutions.

B. States Represented

SFRP -Applicants came from 47 states plus Washington D.C. and Puerto Rico. Selectees represent 44 states plus Puerto Rico.

GSRP - Applicants came from 44 states and Puerto Rico. Selectees represent 32 states.

HSAP - Applicants came from thirteen states. Selectees represent nine states.

Total Number of Participants	
SFRP	188
GSRP	109
HSAP	138
TOTAL	435

Degrees Represented			
	SFRP	GSRP	TOTAL
Doctoral	184	1	185
Master's	4	48	52
Bachelor's	0	60	60
TOTAL	188	109	297

SFRP Academic Titles	
Assistant Professor	79
Associate Professor	59
Professor	42
Instructor	3
Chairman	0
Visiting Professor	1
Visiting Assoc. Prof.	0
Research Associate	4
TOTAL	188

Source of Learning About the SRP		
Category	Applicants	Selectees
Applied/participated in prior years	28%	34%
Colleague familiar with SRP	19%	16%
Brochure mailed to institution	23%	17%
Contact with Air Force laboratory	17%	23%
IEEE Spectrum	2%	1%
BIIHE	1%	1%
Other source	10%	8%
TOTAL	100%	100%

APPENDIX B – SRP EVALUATION RESPONSES

1. OVERVIEW

Evaluations were completed and returned to RDL by four groups at the completion of the SRP. The number of respondents in each group is shown below.

Table B-1. Total SRP Evaluations Received

Evaluation Group	Responses
SFRP & GSRPs	275
HSAPs	113
USAF Laboratory Focal Points	84
USAF Laboratory HSAP Mentors	6

All groups indicate unanimous enthusiasm for the SRP experience.

The summarized recommendations for program improvement from both associates and laboratory personnel are listed below:

A. Better preparation on the labs' part prior to associates' arrival (i.e., office space, computer assets, clearly defined scope of work).

B. Faculty Associates suggest higher stipends for SFRP associates.

C. Both HSAP Air Force laboratory mentors and associates would like the summer tour extended from the current 8 weeks to either 10 or 11 weeks; the groups state it takes 4-6 weeks just to get high school students up-to-speed on what's going on at laboratory. (Note: this same argument was used to raise the faculty and graduate student participation time a few years ago.)

2. 1996 USAF LABORATORY FOCAL POINT (LFP) EVALUATION RESPONSES

The summarized results listed below are from the 84 LFP evaluations received.

1. LFP evaluations received and associate preferences:

Table B-2. Air Force LFP Evaluation Responses (By Type)

		How Many Associates Would You Prefer To Get ? (% Response)											
		SFRP				GSRP (w/Univ Professor)				GSRP (w/o Univ Professor)			
Lab	Evals Recv'd	0	1	2	3+	0	1	2	3+	0	1	2	3+
AEDC	0	-	-	-	-	-	-	-	-	-	-	-	-
WHMC	0	-	-	-	-	-	-	-	-	-	-	-	-
AL	7	28	28	28	14	54	14	28	0	86	0	14	0
FJSRL	1	0	100	0	0	100	0	0	0	0	100	0	0
PL	25	40	40	16	4	88	12	0	0	84	12	4	0
RL	5	60	40	0	0	80	10	0	0	100	0	0	0
WL	46	30	43	20	6	78	17	4	0	93	4	2	0
Total	84	32%	50%	13%	5%	80%	11%	6%	0%	73%	23%	4%	0%

LFP Evaluation Summary. The summarized responses, by laboratory, are listed on the following page. LFPs were asked to rate the following questions on a scale from 1 (below average) to 5 (above average).

2. LFPs involved in SRP associate application evaluation process:
    a. Time available for evaluation of applications:
    b. Adequacy of applications for selection process:
3. Value of orientation trips:
4. Length of research tour:
5 a. Benefits of associate's work to laboratory:
    b. Benefits of associate's work to Air Force:
6. a. Enhancement of research qualifications for LFP and staff:
    b. Enhancement of research qualifications for SFRP associate:
    c. Enhancement of research qualifications for GSRP associate:
7. a. Enhancement of knowledge for LFP and staff:
    b. Enhancement of knowledge for SFRP associate:
    c. Enhancement of knowledge for GSRP associate:
8. Value of Air Force and university links:
9. Potential for future collaboration:
10. a. Your working relationship with SFRP:
    b. Your working relationship with GSRP:
11. Expenditure of your time worthwhile:

(Continued on next page)

12. Quality of program literature for associate:
13. a. Quality of RDL's communications with you:
    b. Quality of RDL's communications with associates:
14. Overall assessment of SRP:

Table B-3. Laboratory Focal Point Reponses to above questions

	AEDC	*AL*	*FJSRL*	*PL*	*RL*	*WHMC*	*WL*
# Evals Recv'd	0	7	1	14	5	0	46
Question #							
2	-	86 %	0 %	88 %	80 %	-	85 %
2a	-	4.3	n/a	3.8	4.0	-	3.6
2b	-	4.0	n/a	3.9	4.5	-	4.1
3	-	4.5	n/a	4.3	4.3	-	3.7
4	-	4.1	4.0	4.1	4.2	-	3.9
5a	-	4.3	5.0	4.3	4.6	-	4.4
5b	-	4.5	n/a	4.2	4.6	-	4.3
6a	-	4.5	5.0	4.0	4.4	-	4.3
6b	-	4.3	n/a	4.1	5.0	-	4.4
6c	-	3.7	5.0	3.5	5.0	-	4.3
7a	-	4.7	5.0	4.0	4.4	-	4.3
7b	-	4.3	n/a	4.2	5.0	-	4.4
7c	-	4.0	5.0	3.9	5.0	-	4.3
8	-	4.6	4.0	4.5	4.6	-	4.3
9	-	4.9	5.0	4.4	4.8	-	4.2
10a	-	5.0	n/a	4.6	4.6	-	4.6
10b	-	4.7	5.0	3.9	5.0	-	4.4
11	-	4.6	5.0	4.4	4.8	-	4.4
12	-	4.0	4.0	4.0	4.2	-	3.8
13a	-	3.2	4.0	3.5	3.8	-	3.4
13b	-	3.4	4.0	3.6	4.5	-	3.6
14	-	4.4	5.0	4.4	4.8	-	4.4

3. 1996 SFRP & GSRP EVALUATION RESPONSES

The summarized results listed below are from the 257 SFRP/GSRP evaluations received.

Associates were asked to rate the following questions on a scale from 1 (below average) to 5 (above average) - by Air Force base results and over-all results of the 1996 evaluations are listed after the questions.

1. The match between the laboratories research and your field:
2. Your working relationship with your LFP:
3. Enhancement of your academic qualifications:
4. Enhancement of your research qualifications:
5. Lab readiness for you: LFP, task, plan:
6. Lab readiness for you: equipment, supplies, facilities:
7. Lab resources:
8. Lab research and administrative support:
9. Adequacy of brochure and associate handbook:
10. RDL communications with you:
11. Overall payment procedures:
12. Overall assessment of the SRP:
13. a. Would you apply again?
    b. Will you continue this or related research?
14. Was length of your tour satisfactory?
15. Percentage of associates who experienced difficulties in finding housing:
16. Where did you stay during your SRP tour?
    a. At Home:
    b. With Friend:
    c. On Local Economy:
    d. Base Quarters:
17. Value of orientation visit:
    a. Essential:
    b. Convenient:
    c. Not Worth Cost:
    d. Not Used:

SFRP and GSRP associate's responses are listed in tabular format on the following page.

Table B-4. 1996 SFRP & GSRP Associate Responses to SRP Evaluation

	Arnold	Brooks	Edwards	Eglin	Griffis	Hanscom	Kelly	Kirtland	Lackland	Robins	Tyndall	WPAFB	average
# res	6	48	6	14	31	19	3	32	1	2	10	85	257
1	4.8	4.4	4.6	4.7	4.4	4.9	4.6	4.6	5.0	5.0	4.0	4.7	4.6
2	5.0	4.6	4.1	4.9	4.7	4.7	5.0	4.7	5.0	5.0	4.6	4.8	4.7
3	4.5	4.4	4.0	4.6	4.3	4.2	4.3	4.4	5.0	5.0	4.5	4.3	4.4
4	4.3	4.5	3.8	4.6	4.4	4.4	4.3	4.6	5.0	4.0	4.4	4.5	4.5
5	4.5	4.3	3.3	4.8	4.4	4.5	4.3	4.2	5.0	5.0	3.9	4.4	4.4
6	4.3	4.3	3.7	4.7	4.4	4.5	4.0	3.8	5.0	5.0	3.8	4.2	4.2
7	4.5	4.4	4.2	4.8	4.5	4.3	4.3	4.1	5.0	5.0	4.3	4.3	4.4
8	4.5	4.6	3.0	4.9	4.4	4.3	4.3	4.5	5.0	5.0	4.7	4.5	4.5
9	4.7	4.5	4.7	4.5	4.3	4.5	4.7	4.3	5.0	5.0	4.1	4.5	4.5
10	4.2	4.4	4.7	4.4	4.1	4.1	4.0	4.2	5.0	4.5	3.6	4.4	4.3
11	3.8	4.1	4.5	4.0	3.9	4.1	4.0	4.0	3.0	4.0	3.7	4.0	4.0
12	5.7	4.7	4.3	4.9	4.5	4.9	4.7	4.6	5.0	4.5	4.6	4.5	4.6
Numbers below are percentages													
13a	83	90	83	93	87	75	100	81	100	100	100	86	87
13b	100	89	83	100	94	98	100	94	100	100	100	94	93
14	83	96	100	90	87	80	100	92	100	100	70	84	88
15	17	6	0	33	20	76	33	25	0	100	20	8	39
16a	-	26	17	9	38	23	33	4	-	-	-	30	
16b	100	33	-	40	-	8	-	-	-	-	36	2	
16c	-	41	83	40	62	69	67	96	100	100	64	68	
16d	-	-	-	-	-	-	-	-	-	-	-	0	
17a	-	33	100	17	50	14	67	39	-	50	40	31	35
17b	-	21	-	17	10	14	-	24	-	50	20	16	16
17c	-	-	-	-	10	7	-	-	-	-	-	2	3
17d	100	46	-	66	30	69	33	37	100	-	40	51	46

4. 1996 USAF LABORATORY HSAP MENTOR EVALUATION RESPONSES

Not enough evaluations received (5 total) from Mentors to do useful summary.

5. 1996 HSAP EVALUATION RESPONSES

The summarized results listed below are from the 113 HSAP evaluations received.

HSAP apprentices were asked to rate the following questions on a scale from
1 (below average) to 5 (above average)

1. Your influence on selection of topic/type of work.
2. Working relationship with mentor, other lab scientists.
3. Enhancement of your academic qualifications.
4. Technically challenging work.
5. Lab readiness for you: mentor, task, work plan, equipment.
6. Influence on your career.
7. Increased interest in math/science.
8. Lab research & administrative support.
9. Adequacy of RDL's Apprentice Handbook and administrative materials.
10. Responsiveness of RDL communications.
11. Overall payment procedures.
12. Overall assessment of SRP value to you.
13. Would you apply again next year? Yes (92 %)
14. Will you pursue future studies related to this research? Yes (68 %)
15. Was Tour length satisfactory? Yes (82 %)

	Arnold	Brooks	Edwards	Eglin	Griffiss	Hanscom	Kirtland	Tyndall	WPAFB	Totals
# resp	5	19	7	15	13	2	7	5	40	113
1	2.8	3.3	3.4	3.5	3.4	4.0	3.2	3.6	3.6	3.4
2	4.4	4.6	4.5	4.8	4.6	4.0	4.4	4.0	4.6	4.6
3	4.0	4.2	4.1	4.3	4.5	5.0	4.3	4.6	4.4	4.4
4	3.6	3.9	4.0	4.5	4.2	5.0	4.6	3.8	4.3	4.2
5	4.4	4.1	3.7	4.5	4.1	3.0	3.9	3.6	3.9	4.0
6	3.2	3.6	3.6	4.1	3.8	5.0	3.3	3.8	3.6	3.7
7	2.8	4.1	4.0	3.9	3.9	5.0	3.6	4.0	4.0	3.9
8	3.8	4.1	4.0	4.3	4.0	4.0	4.3	3.8	4.3	4.2
9	4.4	3.6	4.1	4.1	3.5	4.0	3.9	4.0	3.7	3.8
10	4.0	3.8	4.1	3.7	4.1	4.0	3.9	2.4	3.8	3.8
11	4.2	4.2	3.7	3.9	3.8	3.0	3.7	2.6	3.7	3.8
12	4.0	4.5	4.9	4.6	4.6	5.0	4.6	4.2	4.3	4.5
Numbers below are percentages										
13	60%	95%	100%	100%	85%	100%	100%	100%	90%	92%
14	20%	80%	71%	80%	54%	100%	71%	80%	65%	68%
15	100%	70%	71%	100%	100%	50%	86%	60%	80%	82%

Salahuddin Ahmed's report was not available at the time of publication.

MODELING OF ORGANOHALIDE REACTIONS IN AQUEOUS B12 / Ti(III) SYSTEMS

Leslie Buck
Teaching Fellow / PhD Candidate
Department of Civil and Environmental Engineering

Polytechnic University
Six Metrotech Center
Brooklyn, N.Y. 11201

Final Report for:
Graduate Student Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Tyndall Air Force Base, Panama City, Florida

and

Armstrong Laboratory
Environics Directorate (AL/EQC)

August 1996

MODELING OF ORGANOHALIDE REACTIONS
IN AQUEOUS B12 / Ti(III) SYSTEMS

Leslie Buck
Teaching Fellow / PhD Candidate
Department of Civil and Environmental Engineering
Polytechnic University

Abstract

Development of a kinetic model for the reaction of various organohalides with vitamin B12 and titanium citrate in an aqueous headspace system was attempted. The experimental procedure was simplified using a head space analysis, therefore, all data acquired for model calibration were in total mass per vial concentration units. The parent compound, Perchloroethylene (PCE), was observed experimentally to follow the expected reaction pathway through trichloroethylene (TCE) to the dichloroethenes (DCE) to vinyl chloride to ethylene and finally to ethane. In addition to this, it is proposed that reductive beta elimination is also a mechanism of this reaction due the observed production of acetylene and chloroacetylene(1). Beginning with the outer most limbs of this complex web of pathways, i.e., reactions of acetylene or vinyl chloride to ethene, kinetic constants were determined and fixed. Progress was made up to the root parent, PCE. The model simulates the data well and lends further insight into the true nature of the reaction. The proposed mechanism

$$S + B \underset{k_2}{\overset{k_1}{\leftrightarrows}} SB$$

$$SB + TC \xrightarrow{k_3} P + B$$

accounts for a complexation rate, k_1, and a decomplexation rate, k_2, where S is the substrate, B is the concentration of B12, SB is the substrate/B12 complex, TC is the concentration of titanium citrate, P is the product formed and k_3 is the forward reaction rate constant of SB with Ti(III).

Further progress was made into the development of a model for just the aqueous phase reaction. That is, the model itself incorporates a correction factor based upon dimensionless Henry's constants, K_h'. Given the respective K_h' s of each substrate and product, along with the aqueous and gaseous volumes of the reaction vessel, the model can simulate the reaction progress once the kinetic parameters are properly fitted. It appears that a correction factor must be applied to all rate constants except for the forward complexation rate, k_1.

MODELING OF ORGANOHALIDE REACTIONS IN AQUEOUS B12 / Ti(III) SYSTEMS

Leslie Buck

Introduction

Groundwater contamination by PCE and TCE has become an increasing problem due to their popularity as solvents and degreasers, their use in the dry cleaning industry, and their increasing use in the purification and synthesis of "crack" cocaine. Garden City, L.I., N.Y., for example, has had numerous incidences of drinking-water well contamination. Traditional bioremediation techniques have been unsuccessful due to the tendency for PCE and TCE to terminate reduction at a more harmful product, vinyl chloride, under anaerobic conditions. Research with zero-valent iron sorption has shown promising results, reducing these compounds completely to their non-chlorinated, organic products, ethylene and ethane(2). A similar process occurs with aqueous B12/titanium citrate systems. Vitamin B12 has a cobalt center which is a strong nucleophile. A complexation occurs with B12 and the substrate which enables a reaction with the Ti(III) catalyst. The reaction pathways are surprising similar to that of the zero-valent iron. A mathematical model for the kinetics of this reaction has been developed and is shown here.

Methodology

The software, Scientist(3), was used to alleviate the tedious mathematical computations and integrations of the linear ordinary differential equations determined by the reaction mechanisms. The data used for these simulations was obtained via gas chromatography using 160 mL vials with 60 mL of headspace. The GC was calibrated for total mass per vial concentrations and all results are reported as such. Further details of experimental procedure and controls used in the data collection can be found in this paper(1). A detailed view of the overall reaction pathway is shown in Figure 1. Small "sub-pathways" of this were taken and analyzed individually with their corresponding data. Modeling began with the acetylene to ethylene reaction. Ethane production was very small and not considered in this model. Each sub-reaction was modeled with the same basic mechanism,

$$S + B \underset{k_2}{\overset{k_1}{\leftrightarrows}} SB$$

$$SB + TC \xrightarrow{k_3} P + B$$

where S is the substrate or parent compound, B is the vitamin B12, SB is the substrate/B12 complex, TC is the Ti(III), P is the product formed and k_1 , k_2 , and k_3 are the respective kinetic constants. Initial guesses were made for the three unknown kinetic parameters based upon the given data. A Simplex

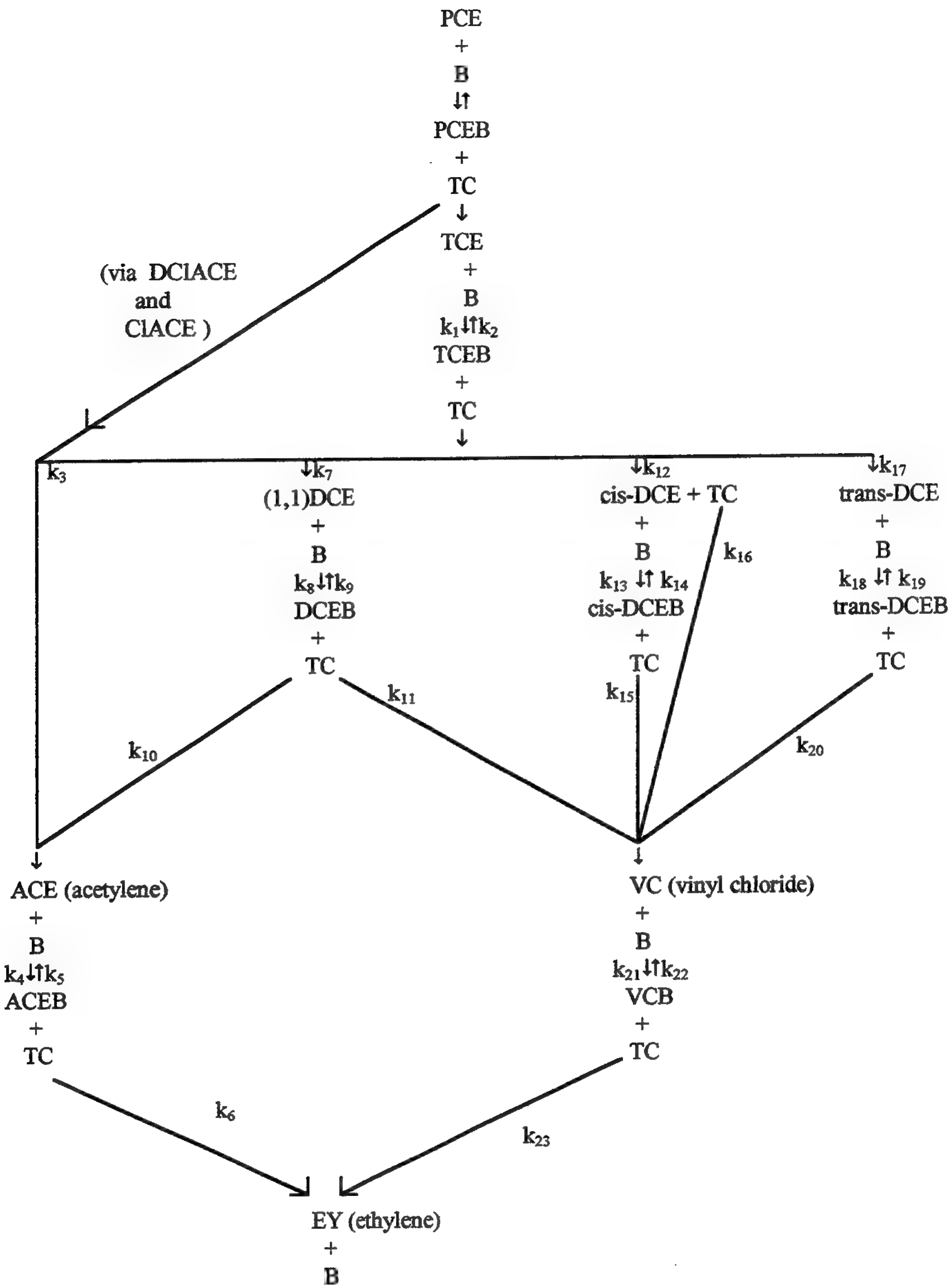


Figure 1 - Pathways indicated by data and used to fit kinetic parameters of model.

smoothing was applied initially, to find a nearby relative minimum if it existed. Then k_2 and k_3 were fixed while a Least Squares fit was performed to best approximate k_1 . k_3 was then fit, followed by k_2 . This was generally the order of decreasing significance. The results of this first data fit of acetylene to ethene follows in Figure 2. To verify the validity of the model, mass balances were performed on each sub-pathway model and on the final overall model from TCE down. Once kinetic rate constants were determined using real data, these parameters were fixed and a simulation was performed. This results of this simulation was used to perform a mass balance on each successive sub-pathway model. The mass balance for acetylene is shown in Figure 3.

Discussion

When mass/vial concentrations were used in data fitting and simulation, the mass balances all checked out perfectly. Data was reported from the GC in mass/vial concentrations. Attempts were made to convert these concentrations to aqueous phase concentrations in mg/l. Using Henry's constants and the known gas and liquid volumes, translation of discrete data points from mass/vial to mg/l was simple. The total mass of the system equals the mass in the aqueous phase plus the mass in the vapor phase or

$$\begin{aligned} M_T &= M_a + M_v \\ &= C_a V_a + C_v V_v \end{aligned}$$

where C_a and C_v are aqueous and vapor phase concentrations and V_a and V_v are aqueous and vapor phase volumes respectively. Solving for C_a and substituting $C_v = K_h'C_a$ where K_h' is the dimensionless Henry's constant for concentration,

$$C_a = \frac{M_T}{V_a + V_v K_h'} = C_F M_T$$

where C_F is the correction factor applied to each discrete data point. This $C_F = (V_a + V_v K_h')^{-1}$ value is a factor of the well known fraction of total mass in the aqueous phase(4),

$$\zeta_a = \frac{1}{1 + K_h'(V_a/V_v)} = \frac{V_a}{V_a + K_h'V_v} = C_F V_a .$$

Initial attempts to incorporate this concentration conversion into the model directly led to erroneous results. Mass balances failed dramatically. The difficulty arose mainly due to the dynamic nature of the gas/liquid partitioning. Simplistic trials of multiplying differential equations in the model uniformly by C_F caused loss and even creation of mass in the mass balance check. The problem becomes apparent when viewed as a simple related rate problem. Again, starting with a mass balance and assuming that any change in volume is insignificant,

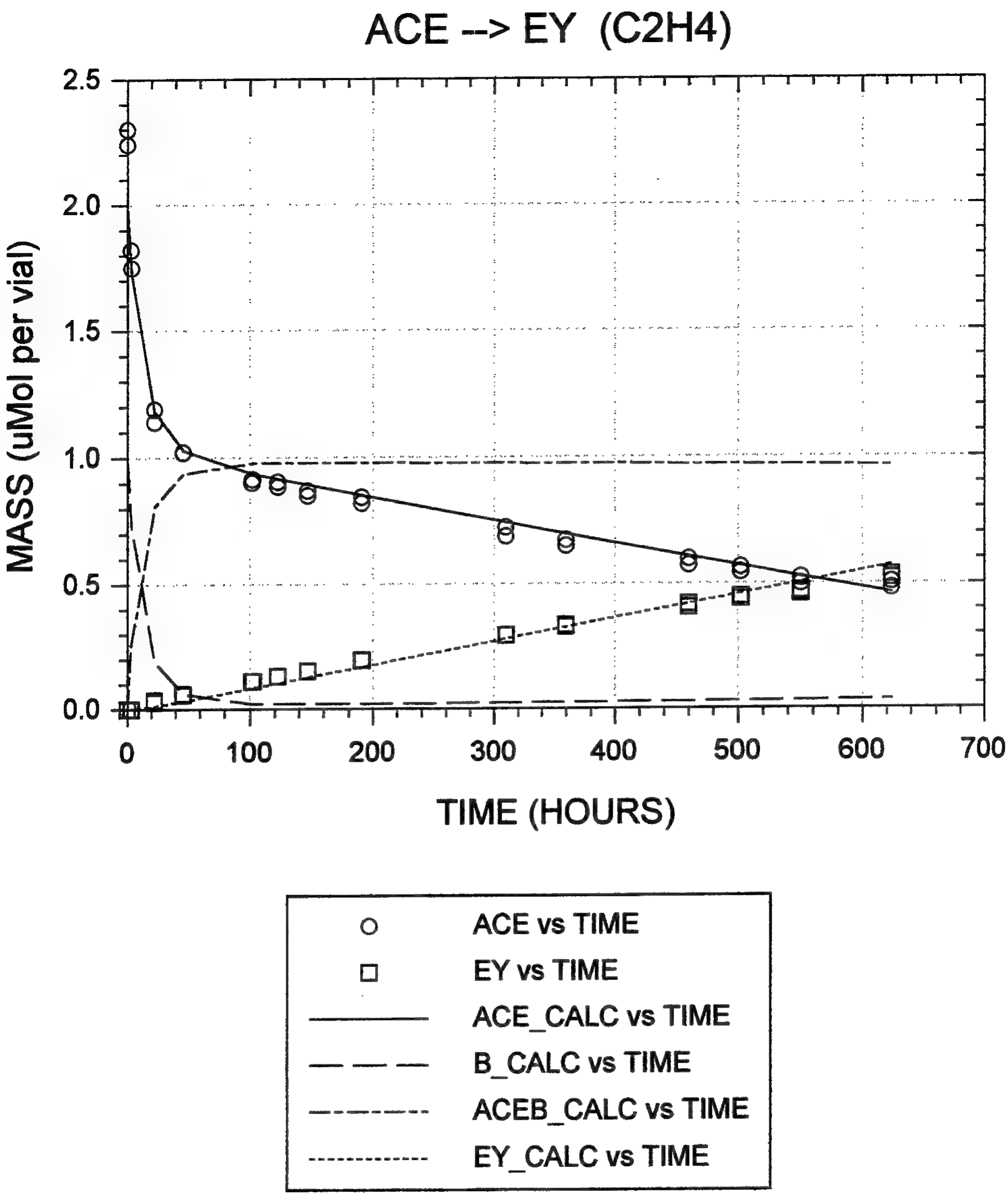


Figure 2 - Acetylene to ethene data as discrete points. Model simulation as solid and dashed lines.

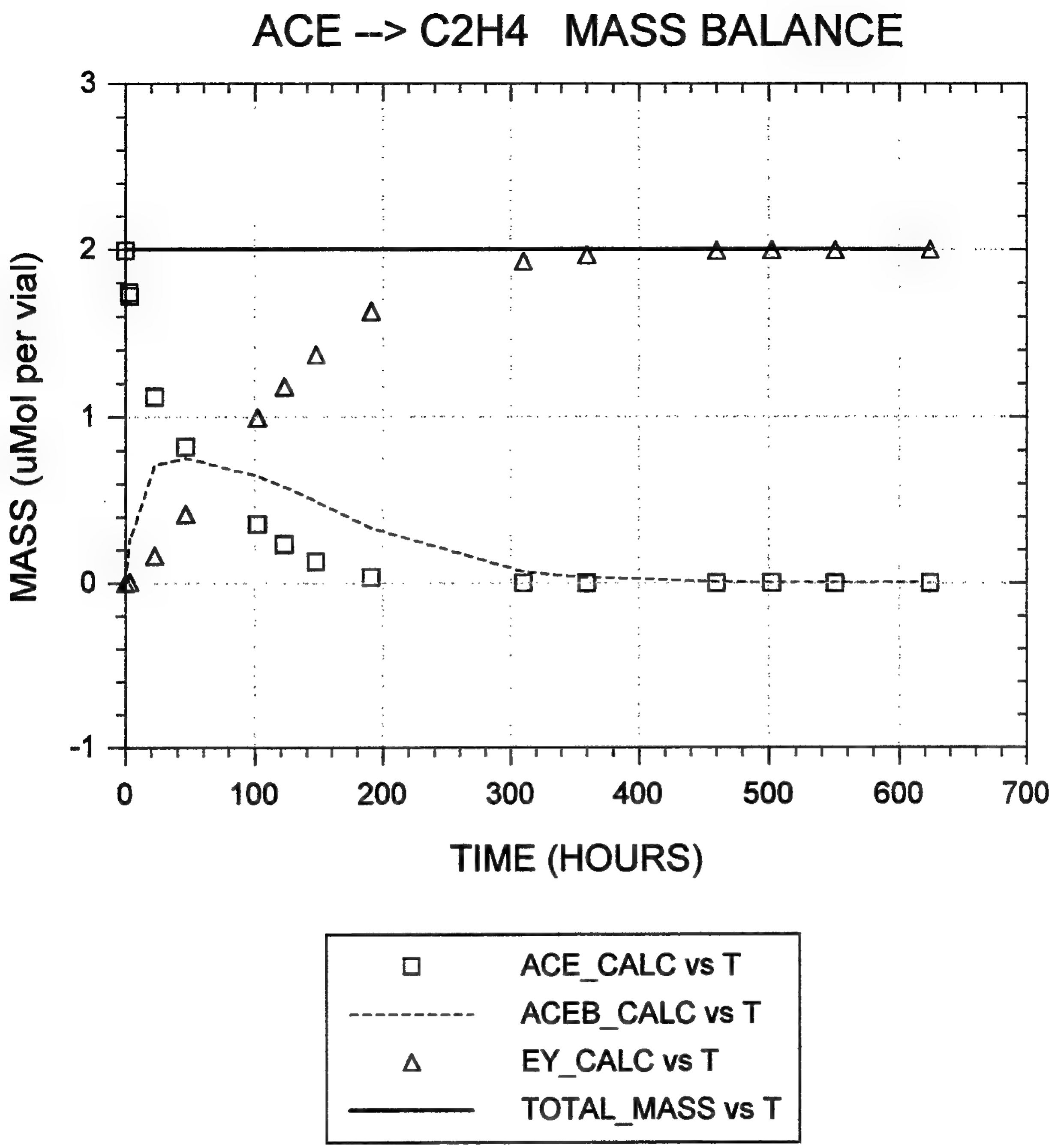


Figure 3 - Mass balance for acetylene to ethene sub-model. Contains only model simulation; no real data.

$$\frac{dM_T}{dt} = V_a\frac{dC_a}{dt} + V_v\frac{dC_v}{dt} \qquad \left(\text{since } \frac{dC_v}{dt} = K_h'\frac{dC_a}{dt}\right)$$

1. $$\frac{dM_T}{dt} = V_a\frac{dC_a}{dt} + V_v Kh'\frac{dC_a}{dt} = (V_a + V_vK_h')\frac{dC_a}{dt} = C_F^{-1}\frac{dC_a}{dt}.$$

Both rates are unknown. Since the complexation mechanism models the total mass/vial data so accurately, a proposed rate of change in mass is,

2. $$\frac{dM_T}{dt} = -k_1 M_TB_T + k_2(MB)_T \quad \text{from } M_T + B_T \underset{k_2}{\overset{k_1}{\leftrightarrows}} (MB)_T.$$

Since B12 and the complex (MB) should only be present in the aqueous phase, it is safe to assume that $B_T = B_a$ and $(MB)_T = (MB)_a$. Finally, making these substitutions and equating equations 1 and 2,

$$\begin{aligned}
\frac{dC_a}{dt} &= C_F[\,-k_1 M_TB_a + k_2(MB)_a\,] \\
&= C_F[\,-k_1(C_aV_a + C_vV_v)B_a + k_2(MB)_a\,] \\
&= C_F[\,-k_1(Va + V_vK_h')C_aB_a + k_2(MB)_a\,] \\
&= C_F[\,-k_1 C_F^{-1}\,CaBa + k_2(MB)_a\,] \\
&= -k_1CaBa + C_F\,k2\,(MB)_a\,.
\end{aligned}$$

The correction factor applies only to the decomplexing rate. This analysis applies to a substrate and not a product nor a substrate which is also a product. When the same logic is applied to a product and a substrate/product, the following equations are derived:

$$\frac{dP_a}{dt} = C_Fk_3(SB)_aTC_a \qquad \text{and} \qquad \frac{dSP_a}{dt} = -k_1\,SP_aB_a + C_F\,[\,k_2(SPB)_a + k_3(SB)_aTC_a]$$

where P_a is the product in the aqueous phase, $(SB)_a$ is the parent substrate complex of that product in the aqueous phase, TC_a is the aqueous Ti(III) concentration, SP_a is the dual substrate/product in the aqueous phase and $(SPB)_a$ is the dual substrate/product complex in the aqueous phase. All rate constants must be corrected for except the forward complexing rate constant, k_1. Bear in mind that the correction factors are unique to each substrate, substrate/product and product.

This analysis was applied to the acetylene to ethene model. The model was calibrated with data corrected to concentrations in the aqueous phase (units of mg/l). The graph of this fit along with its corresponding mass balance are shown in Figures 4 and 5. The fit is excellent. The mass balance is correct to one tenth of a mg/l. Over the 600+ hours, the model loses 0.06 mg/l of total mass from the initial concentration. The dimensionless Henry's constants which are input directly into the model, are only accurate to the tenth's place. The constants used were $K_h' = 3.2$ for acetylene and $K_h' = 0.83$ for ethene, which were determined experimentally. This level of precision may have propagated an error in the model causing the loss of 0.06 mg/l over time.

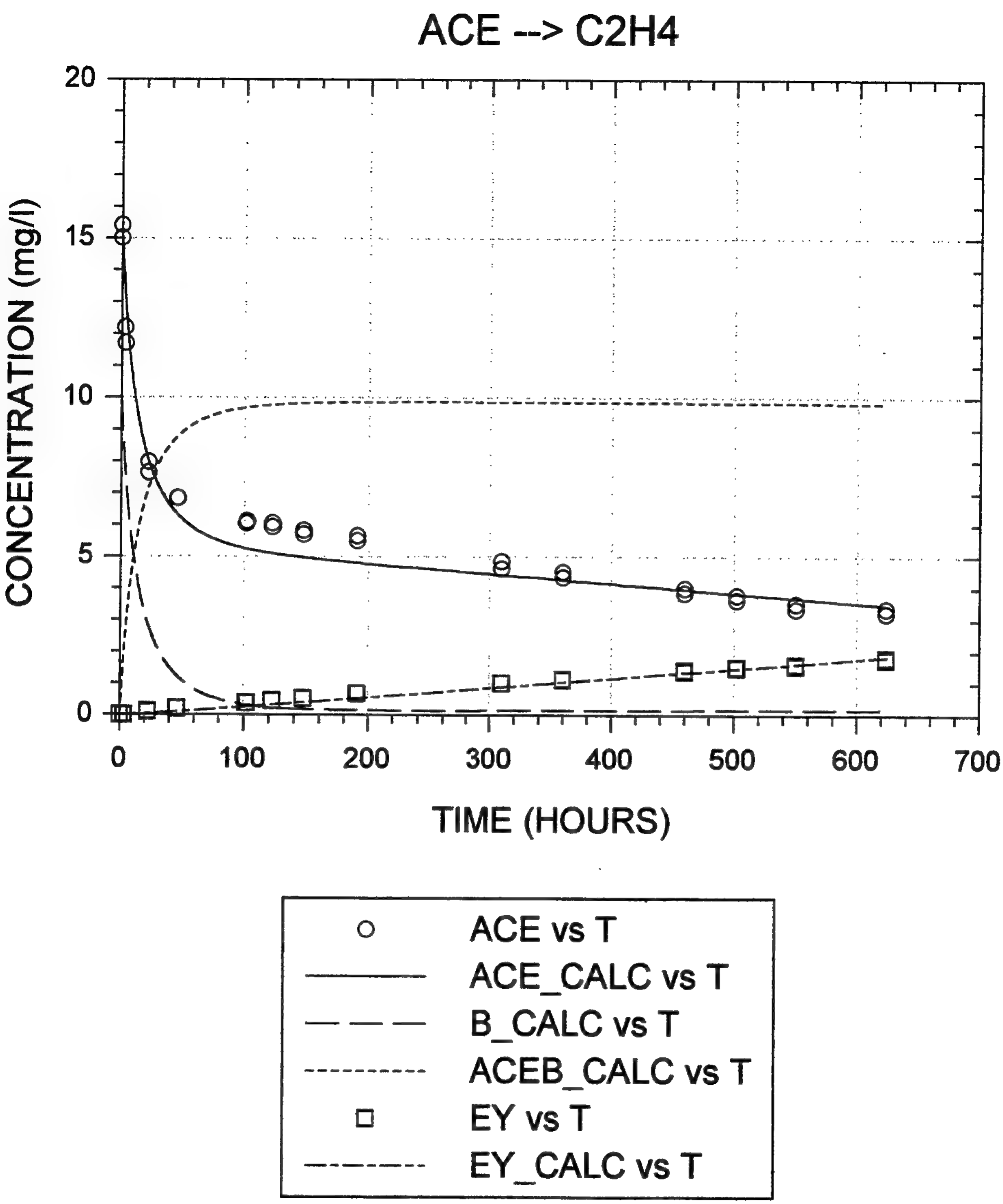


Figure 4 - Acetylene to ethene fit using aqueous-phase model with data corrected to concentration.

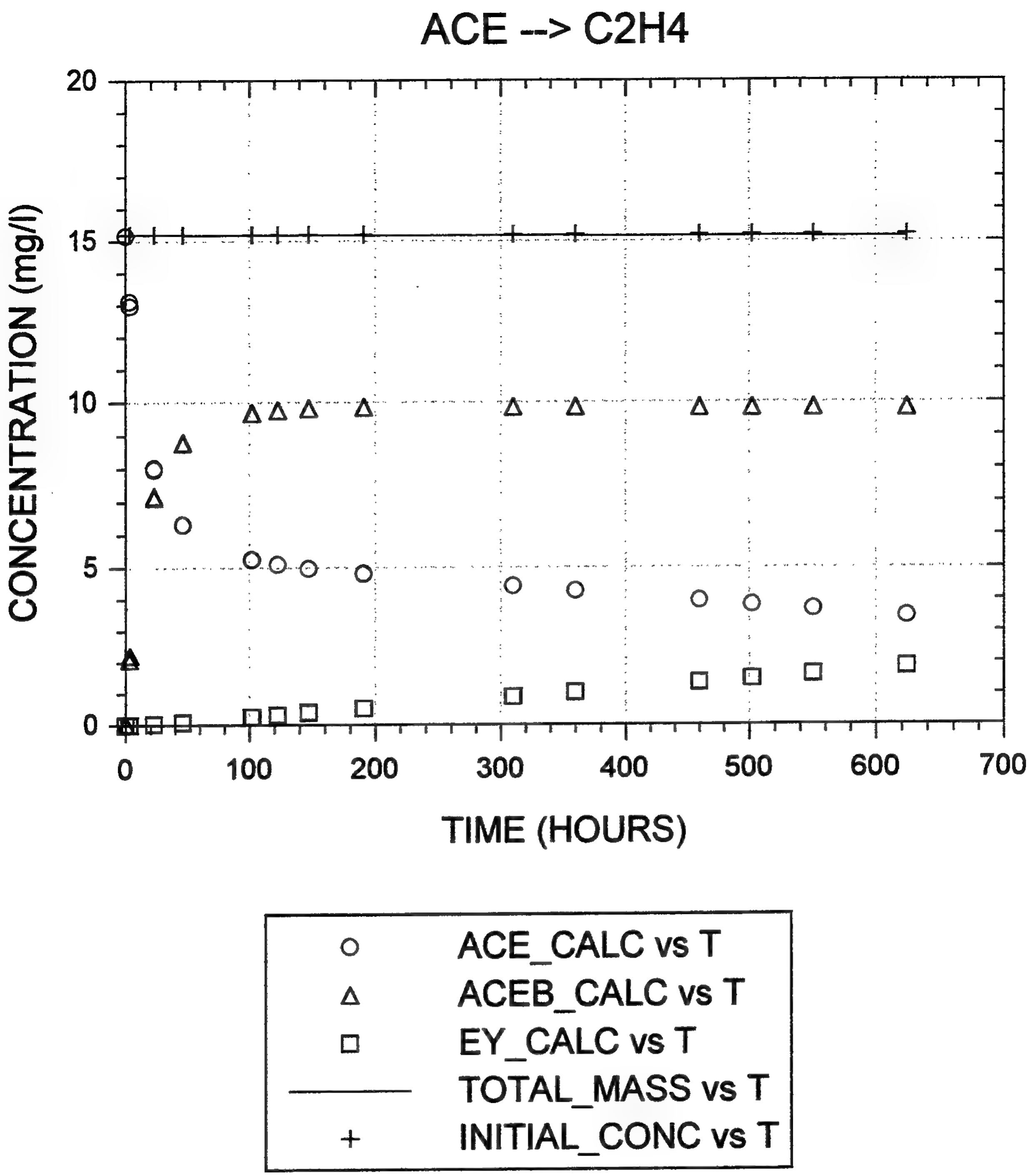


Figure 5 - Mass balance on Figure 4.

Results and conclusion

The original analysis using mass data directly from the GC was carried through to the level of TCE (see Figure 1). These results are illustrated in Figures 6-8. Figure 7 is merely an enlargement of the product region of Figure 6 for clarity. The second analysis, correction for concentration in the aqueous phase only, is still underway. The anticipated outcome is promising, based upon the results for the preliminary sub-pathway model, acetylene to ethene (Figures 4-5). Table 1, below, shows the overall kinetic rate constants for the TCE to ethene model, determined by Scientist(3), which correspond to those diagrammed in Figure 1. It should be noted that k_3 is not the rate constant of TCE to acetylene but rather a representation of two sequential rate constants, one from TCE to chloroacetylene (ClACE) and the other from chloroacetylene to acetylene. ClACE was observed in the experimental analysis but not measured due to the inaccessibility of a standard for that compound.

SUBSTRATE	PRODUCT	PARAMETER	RATE CONSTANT
TCE	----	k_1	0.180503762
"	----	k_2	1.43995242E-34
"	ACE	k_3	0.000152334086
ACE	----	k_4	0.0521951575
"	----	k_5	6.30511676E-23
"	EY	k_6	9.65129077E-6
TCE	DCE	k_7	1.24094955E-5
DCE	----	k_8	0.000186907209
"	----	k_9	2.01429011E-34
"	ACE	k_{10}	0.860091762
"	VC	k_{11}	1.00000000
TCE	CDCE	k_{12}	0.000262366508
CDCE	----	k_{13}	0.00256253974
"	----	k_{14}	1.14001470E-33
"	VC	k_{15}	8.09392535E-7
"	VC	k_{16}	4.81926720E-11
TCE	TDCE	k_{17}	2.71127715E-5
TDCE	----	k_{18}	0.00122050941
"	----	k_{19}	0.000000000
"	VC	k_{20}	4.60811938E-6
VC	----	k_{21}	0.0140521495
"	----	k_{22}	5.13954333E-30
"	EY	k_{23}	3.19230211E-7

Table 1 - Kinetic rate constants for TCE to ethene model.

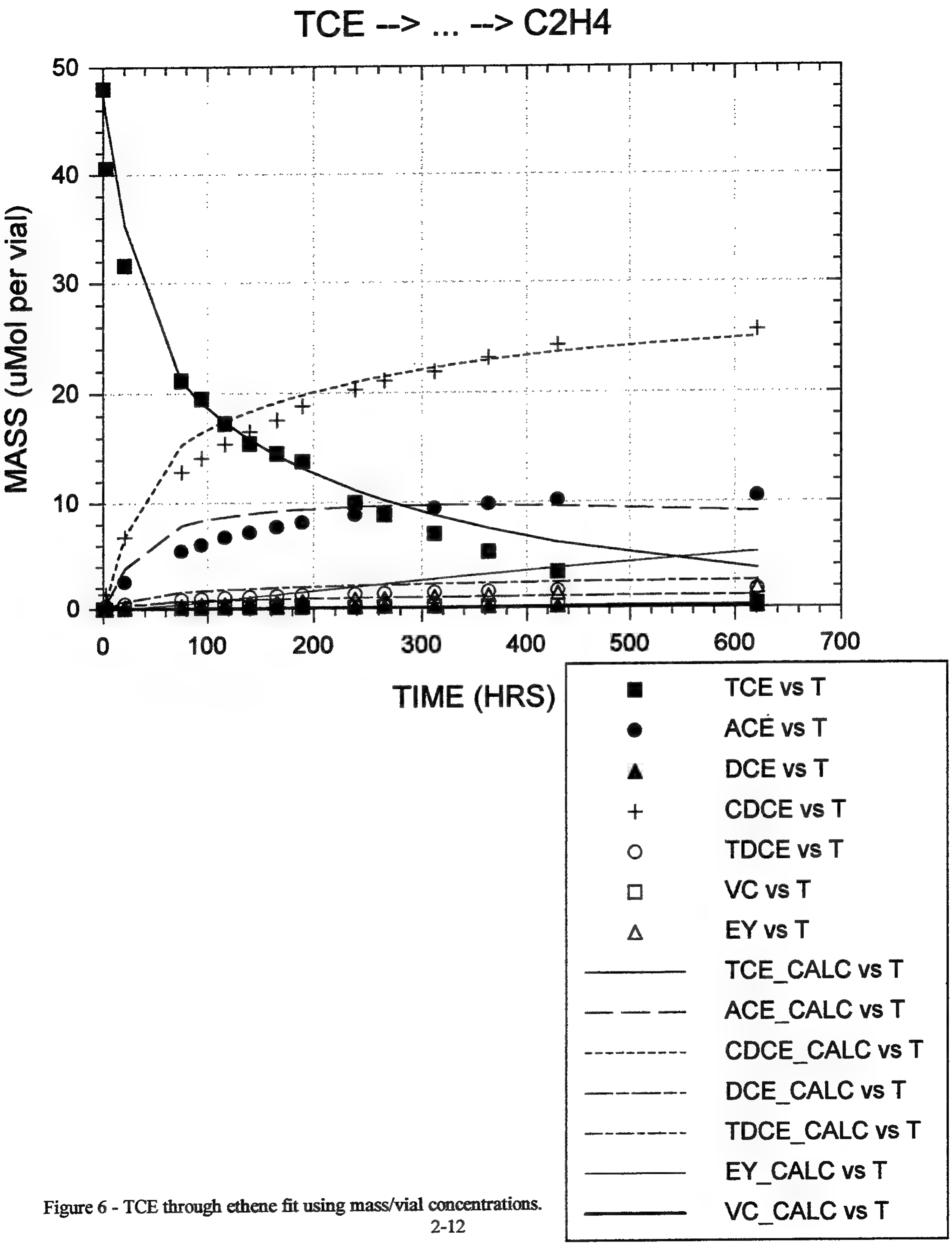


Figure 6 - TCE through ethene fit using mass/vial concentrations.

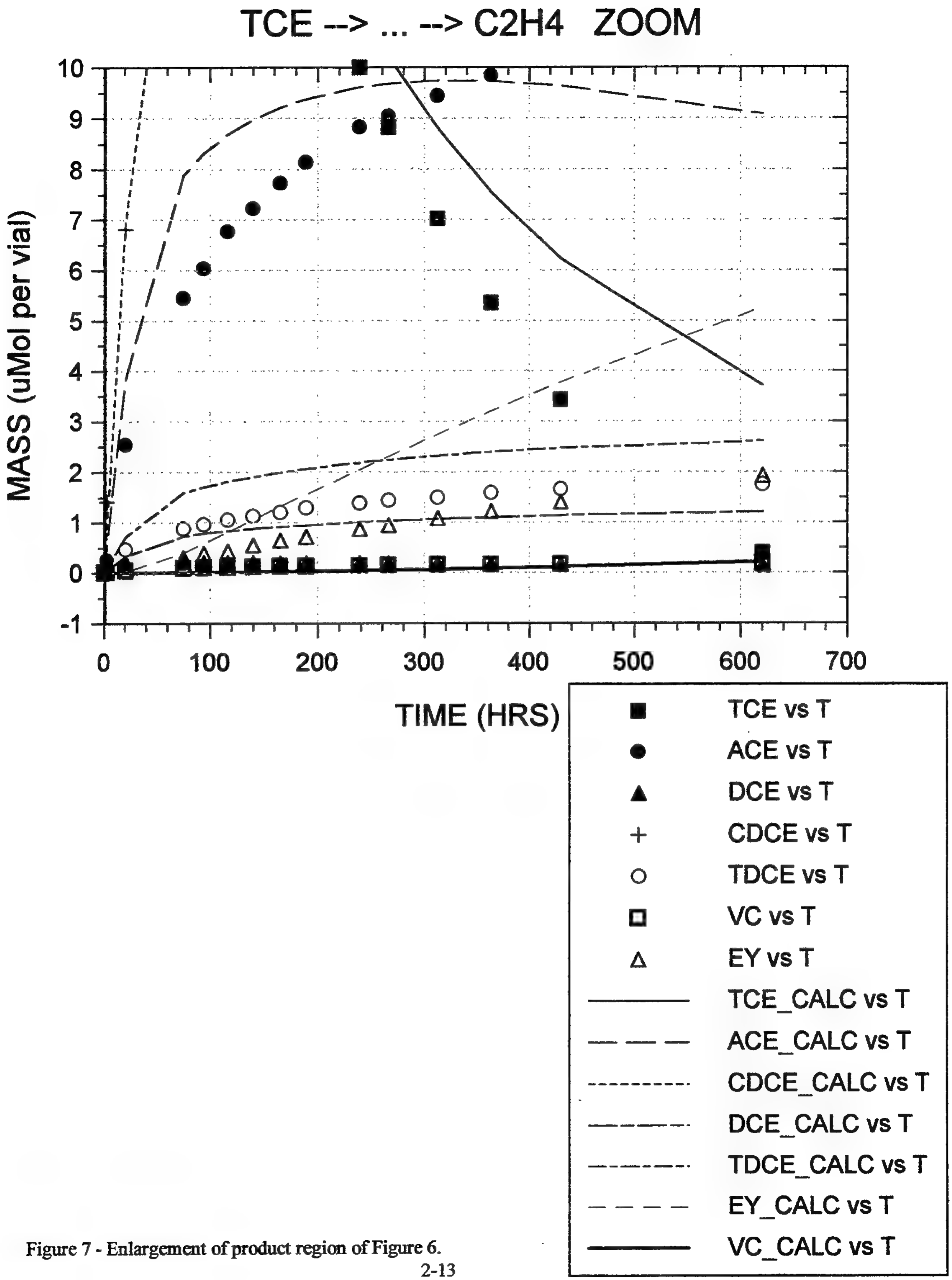


Figure 7 - Enlargement of product region of Figure 6.

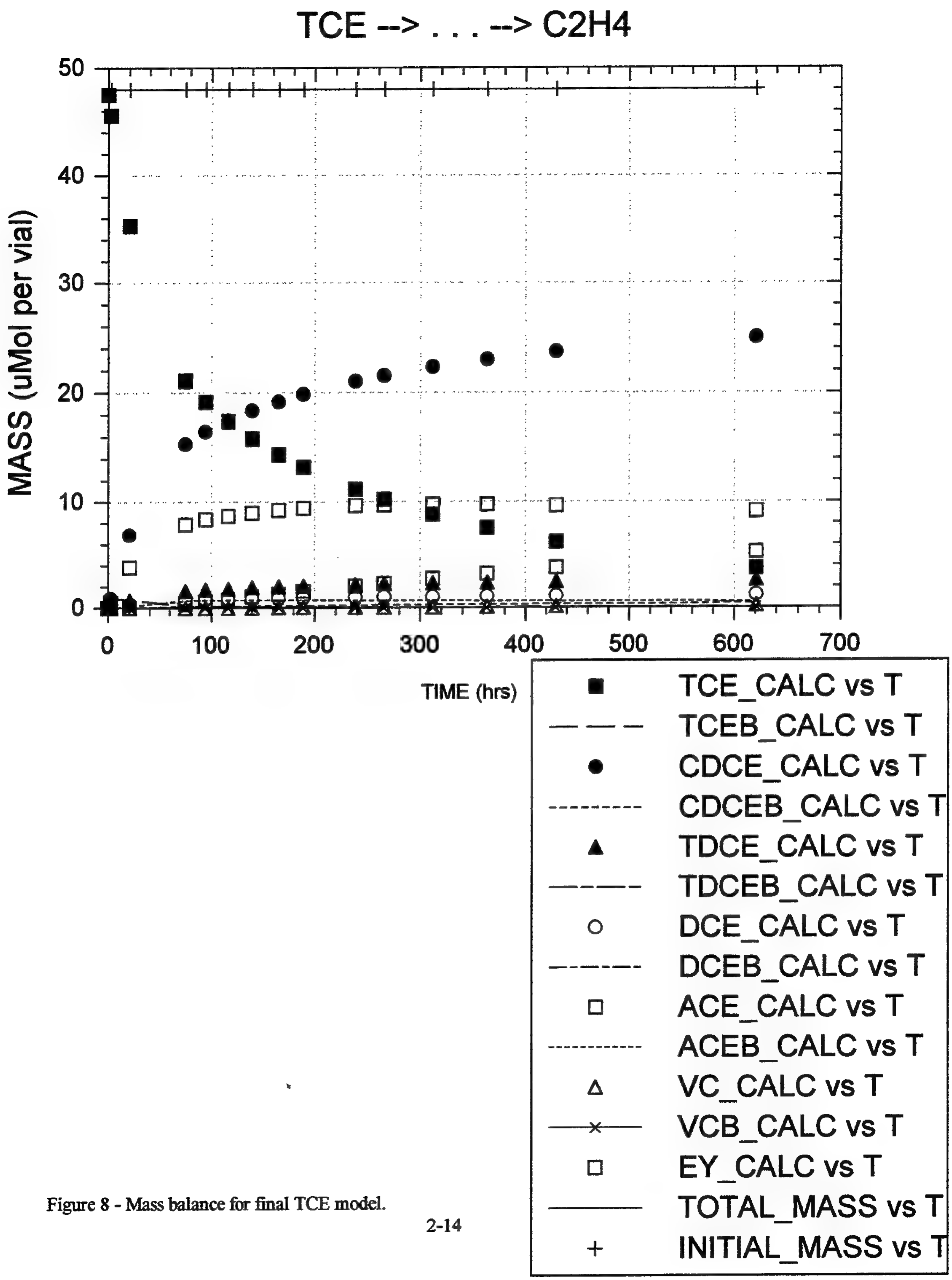


Figure 8 - Mass balance for final TCE model.

References

1. **Burris, David R., Delcomyn, Carrie A., Smith, Mark H. and Roberts, A. Lynn.** *Envir. Sci. Technol.* Revision submitted April 30, **1996.**

2. **Burris, David R., Campbell, Timothy J. and Manoranjan, Valipuram S..** *Envir. Sci. Technol.* **1995**, 29:11:2850-2855.

3. Scientist for Windows, v. 2.01. MicroMath Scientific Software, P.O. Box 21550, Salt Lake City, Utah 84121.

4. **Sawyer, Clair N., McCarty, Perry L. and Parkin, Gene F..** Chemistry for Environmental Engineering, 4th Ed., McGaw-Hill, Inc., **1994** p. 267.

5. **Gossett, James M..** *Envir. Sci. Technol.* **1987,** 21:2:202-208.

6. **Tratnyek, Paul T..** *Chemistry & Industry* **July 1, 1996**, p. 499-503.

Dose-Response of Retinoic Acid-Induced Forelimb Malformations as Determined by Image Analysis

Jerry L. Campbell
Graduate Student
Environmental Health Sciences

University of Georgia
206 Dairy Science Bldg.
Athens, GA 30606

Final Report for:
Graduate Student Research Program
Toxicology Division, Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Armstrong Laboratory
Wright-Patterson Air Force Base, OH

October 1996

Dose-Response of Retinoic Acid-Induced Forelimb Malformations as Determined by Image Analysis

Jerry L. Campbell
Graduate Student
Environmental Health Sciences
University of Georgia

Abstract

Exposure of gestation day 11 mouse embryos to exogenous all-*trans* retinoic acid (RA) results in altered bone development and pattern formation in the limb. The dose-response curve for specific limb malformations remains poorly characterized, a potential impediment to quantitative risk assessment. Therefore, pregnant CD-1 mice were administered a single oral dose of RA (0, 2.5, 10, 30, 60, and 100 mg/kg) on gestation day 11, and day 18 fetuses were examined for forelimb malformations using computerized image analysis. Dose-dependent changes occurred in the size and shape of the scapula, humerus, radius, and ulna, with no effect on the digits. Multiple descriptors of bone size and shape indicate 10 mg/kg to be a near threshold dose for malformations, while 100 mg/kg results in severe alterations in bone size and shape in virtually all forelimbs. By utilizing image analysis to characterize RA-induced forelimb malformations over a broad range, an extremely detailed and highly quantitative analysis of the dose-response relationship was made.

Dose-Response of Retinoic Acid-Induced Forelimb Malformations as Determined by Image Analysis

Jerry L. Campbell

Introduction

In the development of a Biologically Based Dose-Response (BBDR) model for exposure to all-trans-Retinoic Acid (RA), a dose-response curve must first be constructed under the exact experimental conditions used in the model. While some data exists on the incidence and pattern of malformations following near threshold (10 mg/kg maternal body weight) and maximally effective (100 mg/kg maternal body weight) doses of RA, virtually nothing is known about the effects of intermediate doses (Kochhar et al., 1984). Excessive amounts of RA effect almost every organ system when exposure occurs during the appropriate stage of development (Kochhar, 1967 and 1973; Shenefelt, 1972; and Kamm et al., 1984). Exposure of gestation day (GD) 11 mice to RA primarily results in fetal limb and craniofacial malformations (Soprano et al., 1994).

While previous studies have used morphological measurements of the incidences of malformations similar to those described in Kochhar et al. (1984), this research focuses on a more quantitative approach utilizing computerized image analysis to determine the amount of structural changes in fetuses. For this reason, GD 11 CD-1 mice were given a single dose of either vehicle (control), 2.5, 10, 30, 60, or 100 mg of RA per kg maternal body weight. Fetuses were removed on GD 18 and fixed in 95% ethanol. After skeletal regions were stained with alcian blue and alizarin red S, fetuses were stored in a 1:1 solution of glycerin and 70% ethanol until limbs were removed for image analysis. Images of both left and right limbs were saved to disk and analyzed using a Leica Quantimet 570c Image Analysis System. Measurements were taken and correlated with the dose

administered to determine their feasibility for the development a quantitative dose-response curve.

Materials and Methods

Animals and Dosing

Timed-pregnant CD-1 mice were obtained on GD 8 from Charles River Laboratories. GD 0 was determined by the presence of a copulatory plug following mating. Animals were housed individually in polypropylene cages with wood shavings bedding and given laboratory chow (Purina) and water ad libitum. Rooms were kept at a constant temperature with a 12 hour light/dark cycle.

The experiment was carried out in 5 blocks containing 25 to 30 animals. Mice were dosed by oral gavage on the morning of GD 11 with all-trans RA suspended in soybean oil. Doses were adjusted by maternal body weight (determined prior to dosing) to provide 0 (vehicle), 2.5, 10, 30, 60, or 100 mg/kg. Solutions were made by dissolving 10 mg RA in 10 ml of 100% ethanol before serial dilution in oil. Dosing solutions were prepared to ensure that dams of equal weight received equal volumes. Mice remained in home cages until GD 18 when they were sacrificed by CO_2 asphyxiation. Fetuses were given an intra cardiac injection of pentobarbital immediately upon their removal. After weights and crown-rump lengths were determined, fetal heads were removed and sent to NCTR for determination of cleft palate incidence and brain weight. Fetuses were then vicerated and fixed in 95% ethanol for subsequent staining.

Staining Procedure

The staining procedure used was based on a protocol developed by NCTR. Vicerated fetuses were placed in a 70°C water bath for 7 seconds, the skin was subsequently removed, and fetuses were

fixed in 95% ethanol for ≈24 hours. After draining the ethanol, fetuses were placed in a solution containing 150 mg alcian blue in 800 ml of 95% ethanol and 200 ml glacial acetic acid for ≈20 hours in order to stain cartilage. The alcian blue solution was drained and replaced with 95% ethanol for ≈8 hours. Fetuses were then placed in a 0.35% KOH solution overnight to clear tissue. Once the KOH solution was drained, fetuses were submerged in a solution containing 25 g of alizarin red S in 1000 ml of 0.2% KOH for ≈4 hours to stain ossified areas. The alizarin red S solution was then drained and fetuses were placed in a 1:1 solution of 70% ethanol and glycerin. Forelimbs were removed from fetuses prior to imaging by making a cut between the scapula and rib cage using a scalpel (#10).

Image Analysis

In order to capture images, forelimbs were placed in a petri dish containing deionized water and covered with a glass slide. Images were captured using an Olympus dissecting microscope outfitted with a Sony CD-4 digital video camera. Care was taken to maintain identical settings on the microscope and to orient the right and left limbs in the same manner. A Leica Quantimet 570c Image Analysis System was used to save and measure images. Each day, the system was calibrated before limbs were imaged and prior to measuring images saved on a given day, the system was set to the corresponding calibration. Measurements were taken for the scapula and the three long bones (i.e., humerus, radius, and ulna). Widths were determined at both the proximal and distal ends of each bone and lengths were determined to be the distance between the midpoints of the proximal and distal ends of each of the four bones. Area, perimeter, and roundness (a shape measurement) were measured by detecting alizarin red S staining of the ossified regions in each bone. Roundness is based

on a formula including perimeter and area and gives an indication of how close to a perfect circle the bone is (i.e., as a bone shortens or widens becoming more circular in shape, the measurement of roundness will move closer to 1.0, a perfect circle).

Statistics

Significant differences between dose and response were determined for each measurement using analysis of variance (SAS). Simple linear regression (SAS) was used to determine correlation coefficients for individual measurements.

Results

The results used in this report are preliminary. The final database, consisting of measurements for all fetuses imaged, is currently being completed.

For the scapula (Table 1), there appeared to be an increase over the control dose in area, length, width (proximal and distal), and perimeter with the 2.5 and 10 mg/kg doses, although these differences were not significant ($p \leq 0.05$). While there seemed to be a trend for roundness to decrease with increasing doses, the only significant differences occurred between the 100 mg/kg (1.54 ± 0.05) and control (1.45 ± 0.21) doses ($p \leq 0.05$). The results of the scapula did not indicate that there was an increase in response with an increase in dose.

The humerus (Table 2) did indicate a dose-response relationship for several measurements including area ($r=0.978$), length ($r=0.961$), perimeter ($r=0.971$), and roundness ($r=0.930$) with the near threshold doses being 30, 30, 10, and 30 mg/kg, respectively. There was not a significant difference between control, 2.5, and 10 mg/kg doses and between 30 and 60 mg/kg doses for area, length, and roundness ($p \leq 0.05$). For perimeter, there was not a significant difference between control

Table 1. Results of measurements taken for the scapula (average ± standard deviation). For each measurement, results with the same letter are not significantly different ($p \leq 0.05$).

Dose (mg/kg)	# of Fetuses	Area (microns^2 x10^6)	Length (microns)	Proximal width (microns)	Distal width (microns)	Perimeter (microns)	Roundness
0	33	2.35 ± 0.351^{AB}	2084 ± 184^{A}	1756 ± 115^{AB}	681 ± 55^{ABC}	6941 ± 541^{A}	1.54 ± 0.05^{A}
2.5	78	2.47 ± 0.335^{A}	2100 ± 190^{A}	1812 ± 365^{A}	703 ± 67^{A}	7036 ± 502^{A}	1.51 ± 0.04^{AB}
10	62	2.41 ± 0.428^{AB}	2121 ± 157^{A}	1804 ± 103^{A}	699 ± 50^{A}	6957 ± 978^{A}	1.50 ± 0.22^{AB}
30	86	2.34 ± 0.233^{AB}	2019 ± 126^{B}	1773 ± 137^{A}	686 ± 52^{AB}	6846 ± 383^{A}	1.50 ± 0.05^{AB}
60	50	2.26 ± 0.250^{B}	1988 ± 119^{C}	1710 ± 192^{AB}	668 ± 44^{BC}	6690 ± 367^{A}	1.48 ± 0.05^{AB}
100	114	2.10 ± 0.367^{C}	1902 ± 151^{D}	1659 ± 197^{B}	658 ± 49^{C}	6300 ± 962^{B}	1.45 ± 0.21^{B}

Table 2. Results of measurements taken for the humerus (average ± standard deviation). For each measurement, results with the same letter are not significantly different ($p \leq 0.05$).

Dose (mg/kg)	# of Fetuses	Area (microns² x10⁶)	Length (microns)	Proximal width (microns)	Distal width (microns)	Perimeter (microns)	Roundness
0	33	1.75 ± 0.206^{A}	2431 ± 180^{A}	784 ± 51^{A}	604 ± 79^{A}	6997 ± 545^{A}	2.09 ± 0.11^{A}
2.5	78	1.75 ± 0.219^{A}	2401 ± 195^{A}	775 ± 58^{A}	584 ± 101^{AB}	6829 ± 554^{AB}	2.01 ± 0.16^{A}
10	62	1.71 ± 0.296^{A}	2334 ± 155^{A}	780 ± 46^{A}	586 ± 72^{AB}	6596 ± 955^{B}	1.91 ± 0.27^{A}
30	86	1.51 ± 0.241^{B}	1925 ± 249^{B}	785 ± 62^{A}	572 ± 66^{AB}	5913 ± 474^{C}	1.72 ± 0.10^{B}
60	50	1.47 ± 0.288^{B}	1965 ± 329^{B}	759 ± 161^{A}	591 ± 78^{A}	5875 ± 749^{C}	1.77 ± 0.18^{B}
100	114	1.07 ± 0.386^{C}	1438 ± 510^{C}	694 ± 96^{B}	553 ± 63^{B}	4497 ± 1060^{D}	1.49 ± 0.29^{C}

and 2.5 mg/kg, 2.5 and 10 mg/kg, and 30 and 60 mg/kg doses ($p \leq 0.05$). For perimeter the 100 mg/kg dose was significantly different ($p \leq 0.05$) from all other doses. The measurements for proximal width and distal width gave significant differences between the control and 100 mg/kg doses ($p \leq 0.05$) only.

Quantitative measurements of the radius (Table 3) showed a correlation between dose and response for length ($r=0.976$), proximal width ($r=0.970$), perimeter ($r=0.968$), and roundness ($r=0.947$). Near threshold doses were 30, 10, 30, and 10 mg/kg, respectively. Area measurements increased significantly ($p \leq 0.05$) over the control dose for the 10, 30, and 60 mg/kg doses. There was not a significant difference ($p \leq 0.05$) in length between the control, 2.5, and 10 mg/kg doses and between 30 and 60 mg/kg doses. The 100 mg/kg was significantly different from all other doses for length ($p \leq 0.05$). This was the same response given by perimeter. For proximal width, no significant differences ($p \leq 0.05$) occurred between the control and 2.5 mg/kg doses, the 2.5 and 10 mg/kg doses, and the 10, 30, and 60 mg/kg doses. Distal width measurements showed differences that were significant ($p \leq 0.05$) between doses of 10 mg/kg or less and 30 mg/kg or greater. Measurements for roundness had no significant differences ($p \leq 0.05$) between the control and 2.5 mg/kg doses and between the 30 and 60 mg/kg doses.

Table 4 gives the results for measurements of the ulna. As was seen with the radius, there appeared to be a slight increase in area over the control dose for the 2.5, 10, 30, and 60 mg/kg doses, although these increases were not significant ($p \leq 0.05$). The only significant difference ($p \leq 0.05$) for area was with the 100 mg/kg dose. This same pattern was seen with measurements for proximal width. For distal width, there was not a significant difference ($p \leq 0.05$) between control, 2.5, and 10 mg/kg doses and between the 30, 60, and 100 mg/kg doses. Also, there was not a significant

Table 3. Results of measurements taken for the radius (average ± standard deviation). For each measurement, results with the same letter are not significantly different ($p \leq 0.05$).

Dose (mg/kg)	# of Fetuses	Area (microns^2 x10^5)	Length (microns)	Proximal width (microns)	Distal width (microns)	Perimeter (microns)	Roundness
0	33	7.95 ± 0.919^{A}	2247 ± 171^{A}	314 ± 35^{A}	316 ± 48^{A}	5416 ± 402^{A}	2.76 ± 0.16^{A}
2.5	78	8.47 ± 1.06^{AB}	2275 ± 189^{A}	331 ± 37^{AB}	334 ± 42^{A}	5515 ± 460^{A}	2.68 ± 0.18^{A}
10	62	9.48 ± 2.75^{C}	2262 ± 138^{A}	359 ± 36^{BC}	352 ± 65^{A}	5479 ± 820^{A}	2.41 ± 0.36^{B}
30	86	9.33 ± 1.71^{C}	1965 ± 222^{B}	380 ± 50^{C}	401 ± 79^{B}	5040 ± 459^{B}	2.08 ± 0.29^{C}
60	50	9.13 ± 1.35^{BC}	1924 ± 295^{B}	395 ± 71^{C}	412 ± 81^{B}	4944 ± 467^{B}	2.03 ± 0.32^{C}
100	114	7.85 ± 1.91^{A}	1529 ± 497^{C}	460 ± 155^{D}	442 ± 148^{B}	4042 ± 941^{C}	1.64 ± 0.46^{D}

Table 4. Results of measurements taken for the ulna (average ± standard deviation). For each measurement, results with the same letter are not significantly different ($p \leq 0.05$).

Dose (mg/kg)	# of Fetuses	Area (microns^2 x10^6)	Length (microns)	Proximal width (microns)	Distal width (microns)	Perimeter (microns)	Roundness
0	33	1.18 ± 0.147^{A}	2704 ± 205^{A}	652 ± 53^{A}	352 ± 30^{A}	6530 ± 459^{A}	721 ± 4126^{A}
2.5	78	1.25 ± 0.175^{A}	2705 ± 208^{A}	657 ± 50^{A}	374 ± 35^{A}	6617 ± 492^{A}	1891 ± 6113^{A}
10	62	1.27 ± 0.208^{A}	2671 ± 169^{A}	669 ± 46^{A}	388 ± 34^{AB}	6465 ± 942^{A}	2168 ± 6202^{A}
30	86	1.24 ± 0.172^{A}	2281 ± 350^{B}	641 ± 95^{A}	438 ± 82^{C}	5861 ± 609^{B}	2743 ± 6897^{A}
60	50	1.22 ± 0.167^{A}	2198 ± 418^{B}	664 ± 72^{A}	463 ± 86^{C}	5779 ± 811^{B}	1138 ± 4555^{A}
100	114	0.861 ± 0.379^{B}	1634 ± 784^{C}	550 ± 200^{B}	421 ± 156^{BC}	4300 ± 1413^{C}	980 ± 4433^{A}

difference ($p \leq 0.05$) between the 10 and 100 mg/kg doses. Similar results were seen for length and perimeter measurements which had no significant differences ($p \leq 0.05$) between control, 2.5, and 10 mg/kg doses, as well as, between 30 and 60 mg/kg doses. There were no significant differences ($p \leq 0.05$) between the doses for roundness. There was a correlation between dose and response for both length ($r=0.982$) and perimeter ($r=0.966$) measurements.

Discussion

Based on these preliminary results, there was a correlation between dose and response for several measurements. These include perimeter (humerus), proximal width (radius), and roundness (radius). For each of these measurements, the near threshold dose was determined to be 10 mg/kg. The correlation coefficient for each of these measurements ranged from 0.947 to 0.971 indicating that any of these measurements would be useful in a BBDR model. Several other measurements gave 30 mg/kg as the near threshold dose including area, length, and roundness of the humerus along with length and perimeter of the radius and ulna. The correlation coefficient for these measurements ranged from 0.930 to 0.982. While these measurements would be useful in a BBDR model, they appear to be less sensitive to dose than those which determined 10 mg/kg to be the near threshold dose.

Conclusions

The results from this study indicate that quantitative image analysis is a viable tool for the measurement of fetal bone malformations induced by RA. The major benefit of this method is that it relies on a purely quantitative analysis of the limbs to determine malformations. This is a stark

contrast to previous methods which depended on purely subjective morphological determinations of malformations in fetal limbs exposed to RA. This limitation, unfortunately, does not allow results from morphological studies to be compared to mechanistic action, which is neccessary for the creation of a BBDR model. Because of the objective quantitative nature provided by image analysis, it appears that results from this dose-response study can be related to the mechanistic action of RA to provide a basis for the develpment of a BBDR model.

References

Kochhar, D.M. (1967) Teratogenic activity of retinoic acid. *Acta Pathol Microbiol Scand* 70, 398-404.

Kochhar, D.M. (1973) Limb development in mouse embryos. *Teratology* 7, 289-298.

Kochhar, D.M., Penner, J.D., and Hickey, T. (1984) Retinoic acid enhances the displacement of newly synthesized hyaluronate from cell layer to culture medium during early phases of chondrogenesis. *Cell Differ* 14, 213-221.

Kamm, J.J., Ashenfelter, K.O., and Ehmann, C.W. (1984) Preclinical and clinical toxicology of selected retinoids. In: *The Retinoids*. M.B. Sporn, A.B. Roberts, and D.S. Goodman, eds. Academic Press, Orlando, FL. Vol 2, 287-326.

Shenefelt, R.E. (1972) Morphogenesis of malformations in hamsters caused by retinoic acid: Relation to dose and stage of treatment. *Teratology* 5, 103-118.

Soprano, D.R., Gyda, M., Jiang, H., Harnish, D.C., Ugen, K., Satre, M., Chen, L., Soprano, K.J., and Kochhar, D.M. (1994) A sustained elevation in retinoic acid receptor-β2 mRNA and protein occurs during retinoic acid-induced fetal dysmorphogenesis. *Mech Dev* 45, 243-253.

No report was not available at the time of publication.

THE N=2 ANALYTIC SOLUTION FOR THE EXTENDED NONLINEAR SCHRODINGER EQUATION

Julie C. Cwikla
Graduate Student
Department of Mathematics

New York University
Courant Institute of Mathematical Sciences
251 Mercer St.
New York, NY 10012-1185

Final Report for:
Graduate Student Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Armstrong Laboratory

August 1996

THE N=2 ANALYTIC SOLUTION FOR THE EXTENDED NONLINEAR SCHRODINGER EQUATION

Julie C. Cwikla
Graduate Student
Department of Mathematics
New York University
Courant Institute of Mathematical Sciences

Abstract

The extended nonlinear Schrodinger equation (ENLS) and nonlinear optics was studied. To obtain the analytical solution for the ENLS, an understanding of the inverse scattering method was required; as well as a working knowledge of soliton behavior. The analytical solution for the N=2 case for the extended nonlinear Schrodinger equation (ENLS) was derived. The ENLS takes into account the higher order dispersion terms and an analytic solution will aid in particular those researching nonlinear fiber optics.

THE N=2 ANALYTIC SOLUTION TO THE EXTENDED NONLINEAR SCHRODINGER EQUATION

Julie C. Cwikla

Introduction

Nonlinear optics is the study of the interaction of intense laser light with matter.[1] This phenomena, nonlinear optics, is a consequence of modifying optical properties of a material system by the presence of light. Nonlinear optics and solitons have become the nucleus of current optics research studied by mathematicians and physicists alike; and is the focus of this article. Recently nonlinear optics has turned to the behavior of solitons in nonlinear media. Researchers have been using the nonlinear Schrodinger equation for many years. The last decade has seen the emergence of the extended nonlinear Schrodinger equation (ENLS), and specific attention has been given to the higher order components. Until recently, the higher order dispersion components have been casually ignored. Before now, researchers had yet to find an analytical solution to the N=2, second order soliton solution for the ENLS. The following will detail solitons and their effectiveness, the ENLS, the inverse scattering transform and the derivation of the N=2 analytical solution for the ENLS.

Solitons

Solitons are special kinds of waves that can propagate undistorted over long distances and remain unaffected after collision with each other.[2] Picosecond solitons governed by the nonlinear Schrodinger equation (NLS), are being considered as promising elements in communication systems spanning extremely long distances such as transoceanic cables equipped with optical amplifiers.[3] General soliton solutions can be obtained from the wave-envelope equation using the inverse scattering method.[4] Under the right conditions, self-phase modulation

and dispersion cancel each other so that we obtain a pulse that can travel in a nonlinear, dispersive fiber without changing its shape.[5]

The solitons do change *during* a collision. Their speed changes during overlap and there is a tendency to attract each other as they come closer. After a collision, the two pulses separate and move apart without losing their individual soliton nature and without any change in their appearance.[6] This behavior is normal for pulses in linear media, but in nonlinear media one might expect complications since each pulse influences the other by cross-phase modulation.[3] This property is what makes solitons so unique and useful in long range communication.

To help the reader more fully understand the idea of a soliton, included is the following quote by J. Scott Russell from 1834, describing his observations of a soliton.[7]

> *"I was observing the motion of a boat which was rapidly drawn along a narrow channel by a pair of horses, when the boat suddenly stopped - not so the mass of the water in the channel which it had put in motion; it accumulated round the prow of the vessel in a state of violent agitation, then suddenly leaving it behind, rolled forward with great velocity, assuming the form of a large solitary elevation, a rounded, smooth and well-defined heap of water, which continued its course along the channel apparently without change of form or diminution of speed. I followed it on horseback, and overtook it still rolling on at a rate of some eight or nine miles an hour, preserving its original figure some thirty feet long and a foot a foot and a half in height. Its height gradually diminished, and after a chase of one or two miles I lost it in the windings of the channel. Such, in the month of August 1834, was my first chance interview with that rare and beautiful phenomenon which I have called the wave of Translation . . ."*

Russell did extensive experiments in a laboratory scale wave tank and the study of Solitons began. This issue of soliton waves and their existence in water, was finally resolved by Kortweg and de Vries in 1895, but now in the late 1900's, the study of solitons as applied to optics still demands attention and comprehension.

Higher Order Solitons

The higher order solitons governed by the NLS have been studied extensively by M. J. Potasek and A. E. Paul. For a soliton of N order, the input pulse peak power is exactly N^2 times as high as the fundamental (N=1) soliton. The pulse contracts and spreads out to assume initial shape after propagating one period. The contracting and spreading of the second-order soliton repeats itself indefinitely after each soliton period. Fiber losses reduce the energy of the pulse. If the losses are sufficiently small so that the pulse changes only slightly over one soliton period, the pulse can adjust its width to conform to the diminished energy.[6] If the pulse is amplified gradually, it can restore its original shape. If the loss or gain are so high that the soliton changes its energy considerably within one soliton period, it cannot adjust itself and is destroyed.[8]

Schrodinger Equation

Recently experimental progress has been made in obtaining ultrashort femtosecond optical pulses. But as pulse duration shortens, the nonlinear Schrodinger equation is no longer valid and a new approach is required. The extended nonlinear Schrodinger equation is a combination of the nonlinear Schrodinger and a form of the Kortewedg de-Vries equation as follows respectively.

$$iq_t + q_{xx} \pm 2|q|^2 q = 0$$

$$q_t + q_{xxx} \pm 6|q|^2 q_x = 0$$

A multiple-scale perturbation calculation has led to the extended nonlinear Schrodinger equation (ENLS) given by [9]

$$iq_z - \frac{1}{2}\beta_2 q_{tt} + \frac{n_2\omega_o}{c}|q|^2 q - \frac{1}{6}i\beta_3 q_{ttt} + i\frac{2n_2}{c}\left(|q|^2 q\right)_t + i\frac{2n_2\omega_o\gamma}{bc} q\left(|q|^2\right)_t = 0$$

$\beta_3 = \beta_{\omega\omega\omega}$ is the third order derivative of the propagation constant evaluated at ω_o, b is the radius of the frequency dependent electromagnetic mode propagating in the fiber, and γ is a parameter dependent on the fiber geometry. Even though there are many approximations used in deriving the ENLS, there are many applications when these approximations are altogether valid.[6] This equation is related to the inverse scattering method and is used to determine the potential of a system based on its scattering data.

Inverse Scattering Transform

The scatting problem begins with constructing a matrix which contains information about the coefficients and their properties, but the information refers to the asymptotic properties of the wave. This matrix is formed by combining the four waves, from the left and right, each traveling to positive and negative infinity. This matrix reveals information about what is happening at the ends of the waves and from this information we can determine the scattering data, and then perform the inversion. This data consists of the eigenvalues, proportionality constant, transmission coefficients and reflection coefficients. From the scattering data we can recover the potential via the inverse scattering method.[6]

To understand the complex method of inverse scattering we begin by looking at the linear component of the NLS equation.[10] The NLS and the linear portion follow respectively:

$$q_t - i q_{xx} \pm 2 i q^2 q^* = 0 \qquad (1)$$
$$q_t = i q_{xx}$$

Given the complex valued function $q(x,0)$, $-\infty < x < \infty$, with real and imaginary parts which are smooth and decaying at $x = \pm\infty$ so that its Fourier transform

$$\hat{q}(k,0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} q(x,0) e^{-ikx} dx \qquad (2)$$

exists. Now the equation (1) becomes a coupled system of equations in which the evolution of q at some x depends on its neighboring sites. This is an attempt to discretize a continuous system, by reducing a set of partial differential equations to an infinite set of ordinary differential equations. This concept is frequently used in electromagnetic theory.[10] To separate all the components of the equations we use the Fourier transform (2) evaluated at some t.

$$q(x,t) = \int_{-\infty}^{\infty} \hat{q}(k,t) e^{ikx} dk$$

this converts (1) to an easily solved uncoupled set of equations $\hat{q}_t = -i k^2 \hat{q}$ for $\hat{q}(k,t)$ one for each k. So the solution algorithm, most clearly presented by Newell and Maloney in *Nonlinear Optics* is as follows:

$$q(x,0) \xrightarrow{\text{direct transform}} \hat{q}(k,0)$$
$$\downarrow \text{evolve with (1)}$$
$$q(x,t) \xleftarrow{\text{inverse transform}} \hat{q}(k,t)$$

This is also a canonical transformation associated with Hamiltonian mechanics.[10] Equation (1) can be written as

$$q_t = \frac{\delta H}{\delta q^*}, \qquad q_t^* = -\frac{\delta H}{\delta q}, \quad \text{where } H = -i \int_{-\infty}^{\infty} q_x q_x^* \, dx$$

Now, in the Hamiltonian formulation $(q^*, q) = (P, Q)$ where

$$P_t = -\frac{\delta H}{\delta Q}, \qquad Q_t = \frac{\delta H}{\delta P}$$

These equations solve easily by changing to a new set of coordinates $\overline{P}, \overline{Q}$ that depend on P and Q. To preserve the Hamiltonian relationships, allow H to depend only on $\overline{P}$ Then

$$\overline{P} = \overline{P_o}, \qquad \overline{Q} = \left(\frac{\delta H}{\delta \overline{P}}\right)_{\overline{P_o}} t + \overline{Q_o}$$

and everything is solved. In Fourier coordinates,

$$H = -2\pi i \int k^2 \hat{q}(k,t)\, \hat{q}^*(k,t)\, dk$$

Now choose

$$\overline{P}(k,t) = 2\pi\, \hat{q}(k,t)\, \hat{q}^*(k,t), \qquad \overline{Q}(k,t) = i\, Arg\, \hat{q}(k,t) = \frac{1}{2} \tan^{-1} \frac{\hat{q}}{\hat{q}^*}$$

Then $H = -i \int k^2\, \overline{P}(k,t)\, dk, \quad \overline{P}_t = 0, \quad \overline{Q}_t = i \frac{d}{dt} Arg\, \hat{q}(k,t) = -i k^2$

To affirm that (3) is in fact a canonical transformation, use the exterior product (wedge) to prove that

$$\int \delta P \wedge \delta Q \, dx = \int \delta \overline{P} \wedge \delta \overline{Q} \, dk$$

The Fourier transform is a canonical transformation that takes the equations from a highly coupled linear system (1) to a separable one. The inverse scattering transform (IST) plays the same role with nonlinear partial differential equations (p.d.e.); it is a canonical transformation[8] that when applied to these equations separates each into an uncoupled set of equations for the action $(\overline{P})$ and angle $(\overline{Q})$ variables of the natural normal modes of the nonlinear system.[10]

The goal is to reduce a nonlinear p.d.e. to a set of ordinary differential equations(o.d.e.), and if possible, the equation is said to be integrable.

Next, consider the IST for the NLS. Begin by constructing a matrix containing information about the asymptotic properties of the wave.[10] Consider $V = \begin{pmatrix} v_1(x,t) \\ v_2(x,t) \end{pmatrix}$, where

$$V_x = \begin{pmatrix} -i\zeta & q(x,t) \\ r(x,t) & i\zeta \end{pmatrix} V \tag{3}$$

$$V_t = \begin{pmatrix} -2i\zeta^2 - iqr & 2\zeta q + iq_x \\ 2\zeta r - ir_x & 2i\zeta^2 + iqr \end{pmatrix} V + cV \tag{4}$$

The two parameters ζ and c are arbitrary, with c used to normalize a solution. The other parameter ζ is crucial in the theory. The compatibility of (3) and (4) allows the next pivotal step, for an arbitrary ζ,

$$\begin{aligned} q_t &= i(q_{xx} - 2q^2 r) \\ r_t &= -i(r_{xx} - 2qr^2) \end{aligned} \tag{5}$$

Now to solve the NLS proceed just as before using three steps.

$$\begin{array}{ccc} q(x,0),\ r(x,0) & \xrightarrow{\text{direct transform}} & S(t=0) \\ & & \downarrow \text{time evolution} \\ q(x,t),\ r(x,t) & \xleftarrow{\text{inverse transform}} & S(t=t) \end{array}$$

This scheme is analogous to that of the strictly linear case. So the first step is to understand the properties of the scattering data $S(t=0)$ given $q(x,0)$, $r(x,t)$. Next determine the time evolution and finally reconstruct $q(x,t)$, $r(x,t)$ given $S(t=t)$

1. The Direct Transform[10]

Define solutions to (3) as the Jost functions, they have a particular behavior as $x \to \pm\infty$.

Define the solutions:

$$\varphi \to \begin{pmatrix} 1 \\ 0 \end{pmatrix} e^{-i\zeta x} \qquad x \to \infty$$

$$\varphi \to \begin{pmatrix} 0 \\ -1 \end{pmatrix} e^{i\zeta x} \qquad x \to -\infty$$

$$\varphi \to \begin{pmatrix} 0 \\ 1 \end{pmatrix} e^{i\zeta x} \qquad x \to +\infty$$

$$\varphi \to \begin{pmatrix} 1 \\ 0 \end{pmatrix} e^{-i\zeta x} \qquad x \to +\infty$$

It is obvious $\varphi, \bar{\varphi}$ are linearly independent, and their Wronskian[8] $W(\varphi, \bar{\varphi}) = \varphi_1 \bar{\varphi}_2 - \bar{\varphi}_1 \varphi_2$ is constant, and by (6) equals -1. Similarly $W(\psi, \bar{\psi}) = -1$. Since (3) is a second-order system two linearly independent solutions result and the following relationship must hold:

$$\begin{aligned} \varphi(x,t;\zeta) &= a(\zeta,t)\,\bar{\psi}(x,t;\zeta) + b(\zeta,t)\,\psi(x,t;\zeta) \\ \bar{\varphi}(x,t;\zeta) &= \bar{b}(\zeta,t)\,\bar{\psi}(x,t;\zeta) - \bar{a}(\zeta,t)\,\psi(x,t;\zeta) \end{aligned} \tag{7}$$

or

$$\Phi = (\varphi, -\varphi) = \Psi \mathrm{A} = (\bar{\psi}, \psi)\begin{pmatrix} a & -\bar{b} \\ b & \bar{a} \end{pmatrix}.$$

The matrix A is called the "scattering" matrix, since it describes how the fundamental solution matrix, with asymptotic behavior $\begin{pmatrix} e^{-i\zeta x} & 0 \\ 0 & e^{i\zeta x} \end{pmatrix}$ at $x = -\infty$ looks like at $x = +\infty$. Since $W(\varphi, \bar{\varphi}) = W(\psi, \bar{\psi}) = -1$, we have $a\bar{a} + b\bar{b} = 1$.

2. The Time Evolution of the Scattering Data[10]

First note that φ and $\bar{\psi}$ satisfy (4) for $c = +2i\zeta^2$ and $\bar{\varphi}$ and ψ satisfy (4) for

$c = -2i\zeta^2$. Now let $x \to +\infty$ and substitute $\varphi = \begin{pmatrix} a e^{-i\zeta x} \\ b e^{i\zeta x} \end{pmatrix}$ into (4) and get

$$a_t = 0$$

$$b_t = 4i\zeta^2 b$$

Since $a(\zeta, t)$ is not time dependent, either are its zeros ζ_k, $k = 1, \ldots N$ in the upper half

ζ – plane. Using the definition $\varphi(x, \zeta_k) = b_k \psi(x, \zeta_k)$ and differentiating with respect to

time, notice the time dependence of $b_k(t)$. Now, ascertain:

$$\varphi_t(x, \zeta_k) = \begin{pmatrix} -iqr & 2\zeta_k q + iq_x \\ 2\zeta_k r - ir_x & 4i\zeta_k^2 + iqr \end{pmatrix} \varphi(x, \zeta_k)$$

$$= b_{kt} \psi(x, \zeta_k) + \begin{pmatrix} -4i\zeta_k^2 - iqr & 2\zeta_k q + iq_x \\ 2\zeta_k r - ir_x & iqr \end{pmatrix} b_k \psi(x, \zeta_k),$$

which implies

$$b_{kt} = 4i\zeta_k^2 b_k$$

3. The Inverse Transform[10]

Now the final step is to reconstruct $q(x, t)$ and $r(x, t)$ given

$$S\ (a(\zeta),\ b(\zeta),\ \zeta\ real;\ (\zeta_k, b_k)_1^N\ )$$

First construct the solution $\varphi\, e^{-i\zeta x}$ and then use the fact that one of the properties of ζ relates to $q(x)$. One can show that $\lim_{\zeta\to\infty} 2i\zeta\psi_1 e^{-i\zeta x} = q(x)$ (8)

and using this fact generate $q(x,t)$. Consider the function $\left(\varphi(x,\zeta)\, e^{i\zeta x} \Big/ a(\zeta) \right)$ which is meromorphic, analytic except at poles $z=\zeta_k$, in the upper half of the plane. From (7) we obtain

$$\frac{\varphi\, e^{i\zeta x}}{a} = \bar{\psi}\, e^{i\zeta x} + \frac{b}{a}\psi\, e^{i\zeta x} \tag{9}$$

on the real ζ axis. The following is a version of the Riemann-Hilbert problem. The goal is to construct functions $\frac{\varphi\, e^{i\zeta x}}{a}$ meromorphic with a finite number of poles at given locations ζ_k in the upper half plane, tending to $\begin{pmatrix}1\\0\end{pmatrix}$ as $\zeta\to\infty$, and $\bar{\psi}\, e^{i\zeta x}$, analytic in the lower half plane, tending to $\begin{pmatrix}0\\1\end{pmatrix}$ as $\zeta\to\infty$. The $\mathrm{Im}\,\zeta\le 0$, with difference on the real axis separating the domains of meromorphic and analytic behavior. This is given by the following function $\left[\frac{(\zeta,t)}{a(\zeta)\,\psi(x,\zeta)\,e^{i\zeta x}}\right] e^{i\xi x}$. To solve this Riemann-Hilbert problem evaluate

$$\int_{-\infty}^{\infty} \frac{\varphi(x,\zeta')\, e^{i\zeta' x}}{a(\zeta')(\zeta'-\zeta^*)}\, d\zeta' \tag{10}$$

The Riemann-Hilbert problem eventually reduces to an integral equation of Gelfand-Levitan-Marcenko type by taking a Fourier transform. Hence, we arrive at the following equation:

$$\begin{pmatrix}\psi_2^*(x,\zeta)\\ -\psi_1^*(x,\zeta)\end{pmatrix} e^{i\zeta^* x} = \begin{pmatrix}1\\0\end{pmatrix} - \sum_1^N \frac{\gamma_k \Psi(x,\zeta_k)\, e^{i\zeta_k x}}{\zeta_k - \zeta^*} \tag{11}$$

From this equation and (8)

$$q(x,t) = -2i \sum_{1}^{N} \gamma_k^* \psi_2^* (x,\zeta_k) e^{-i\zeta_k x} \tag{12}$$

Now the objective is to find $\psi_1(x,\zeta_k)$, $\psi_2(x,\zeta_k)$, $k = 1 \ldots N$, by setting $\zeta = \zeta_j$ in (11) and deriving 2N equations for the 2N unknowns:

$$u_{lk} = \sqrt{\gamma_k}\, \psi_1 (x,\zeta_k), \qquad l = 1,2; \quad k = 1,\ldots N$$

For convenience set

$$\sqrt{\gamma_k}\, e^{i\zeta_k x} = \lambda_k, \quad k = 1 \ldots N$$

and now (11) can be written as:

$$\begin{aligned} u_1 &= -B u_2^* \\ (I + B^* B) u_2^* &= \lambda^* \\ B &= \left(\frac{\lambda_j \lambda_k^*}{\zeta_j - \zeta_k^*} \right), \end{aligned} \tag{13}$$

with u_1 and u_2 column vectors $(u_{11}, \ldots u_{1N})$ and $(u_{21}, \ldots u_{2N})$. From (12) and (13) we get the desired equation with known variables which will be used in the derivation of the analytical solution of the N=2 case for the extended nonlinear Schrodinger equation:

$$q(x,t) = -2i \sum_{k=1}^{N} \lambda_k^* u_{2k}^* . \tag{14}$$

Derivation of the N=2 Analytic Solution for the ENLS

The scattering data evolves according to the following equations:[7]

1. transmissions coefficients

$$r_{-}(\rho,t) = r_{-}(\rho,0)$$

$$\tilde{r}_{t}(\rho,t) = \tilde{r}_{t}(\rho,0)$$

2. reflection coefficients

$$br(\rho,t) = br(\rho,0)\, e^{(-2a_3\rho^3 + 4i\rho^2)\,t}$$

$$\tilde{b}r(\rho,t) = \tilde{b}r(\rho,0)\, e^{(2a_3\rho^3 - 4i\rho^2)\,t}$$

3. proportionality constants

   for eigenvalues $\rho_k \in C_+,\ \tilde{\rho}_k \in C_-$

$$C_k(t) = C_k(0)\, e^{(-2a_3\rho^3 + 4i\rho^2)\,t}$$

$$\tilde{C}_k(t) = \tilde{C}_k(0)\, e^{(2a_3\rho^3 - 4i\rho^2)\,t}$$

Define

$$\psi_{1k}(x) = \sqrt{C_k}\, e^{i\rho_k x}\, \hat{N}_1(x,\rho_k)$$

$$\psi_{2k}(x) = \sqrt{C_k}\, e^{i\rho_k x}\, (1 + \hat{N}_2(x,\rho_k))$$

$$\lambda_k(x) = \sqrt{C_k}\, e^{i\rho_k x}$$

Then we can derive the initial coupled equations from

$$\psi_{ij} = -\sum_{k=1}^{N} \frac{\lambda_j \overline{\lambda_k}}{\rho_j - \overline{\rho_k}}\, \psi_{2k}$$

$$\psi_{2j} - \lambda_j = \sum_{k=1}^{N} \frac{\lambda_j \overline{\lambda_k}}{\rho_j - \overline{\rho_k}}\, \psi_{1k}$$

Our initial coupled equations are as follows:

let $\psi_{11} = A,\quad \psi_{12} = B,\quad \psi_{21^*} = C,\quad \psi_{22^*} = D$

and $\alpha = \dfrac{\lambda_1 \lambda_1^*}{\rho_1 - \rho_1^*},\quad \delta = \dfrac{\lambda_2 \lambda_1^*}{\rho_1 - \rho_2^*},\quad \varepsilon = \dfrac{\lambda_2 \lambda_1^*}{\rho_2 - \rho_1^*},\quad \theta = \dfrac{\lambda_2 \lambda_2^*}{\rho_2 - \rho_2^*}$

(1) $A + \alpha C + \delta^{*} D = 0$

(2) $\alpha A + \delta B + C = 0$

(3) $B + \varepsilon C + \theta D = 0$

(4) $-\varepsilon^{*} A + \theta B + D = 0$

Next, manipulate the four equations to solve for C and D..

$C = \dfrac{\lambda_1^* + D\,(\theta\varepsilon - \alpha\varepsilon^*)}{1-\alpha^2-\varepsilon^2}$, and $D = \dfrac{\lambda_2^*\,(1-\alpha^2-\varepsilon^2) + \lambda_1^*\,(\theta\varepsilon - \alpha\varepsilon^*)}{(1-\theta^2-(\varepsilon^*)^2)\,(1-\alpha^2-\varepsilon^2) - (\theta\varepsilon - \alpha\varepsilon^*)^2}$

To help consolidate the terms we will substitute $\rho_1 = \xi_1 + i\eta_1 = \frac{i}{2}$ and $\rho_2 = \xi_2 + i\eta_2 = \frac{3}{2}i$,

upon the recommendation of Agrawal.[1] With this substitution we obtain:

$\alpha = -i\left|\lambda_1\right|^2,\ \delta = -\varepsilon,\ \varepsilon = -\frac{1}{2}\,i\lambda_1^*\,\lambda_2,\ \theta = -\frac{1}{3}\,i\left|\lambda_2\right|^2.$

For the extended nonlinear Schrodinger equation, an a_3 term is introduced. This parameter is related to higher order dispersion. In the case of the NLS, this a_3 term was zero and the higher order dispersion ignored. For the ENLS let $A\_\rho = a_3\,\rho^3 + a_2\,\rho^2$ and substitute this into the equation for lambda which will eventually be substituted into q. $\lambda_k = \sqrt{C_k(0)}\,e^{i\rho_k x}$ using the time evolution of C_k. With the substitution of the lambdas we are able to combine like terms and simplify the expression. To verify the equation's validity let the a_3 term representing the higher order dispersion, go to zero. Collate the *exact* solution for the N=2 case[11] and the analytical solution. Using equation (14) $q = -2i\ [\lambda_1^* C + \lambda_2^* D]$ we achieve the desired solution:

$$q = -2i\left[\frac{(\lambda_1^*)^2}{1 + |\lambda_1|^4 + \frac{1}{4}(\lambda_1^*)^2 \lambda_2^2} + D\left[\lambda_2^* - \frac{\lambda_1^*\left(\frac{1}{6}\lambda_1^* \lambda_2 |\lambda_2|^2 + \frac{1}{2}\lambda_1 \lambda_2^* |\lambda_1|^2\right)}{1 + |\lambda_1|^4 + \frac{1}{4}(\lambda_1^*)\lambda_2^2}\right]\right]$$

where D is the same as above. This is the analytic solution for the N=2 soliton case for the ENLS, found in a joint effort with M. J. Potasek, published here for the first time.

Conclusion

Slowly varying envelope assumptions and the use of a terminated Taylor expansion provides an incomplete analysis. When using the nonlinear Schrodinger equation, researchers have not been able to retain all information, particularly the linear portion. As we delve further into experimentation with the ultra-short pulses it is possible the assumptions that once would suffice may prove inadequate. The physical evidence thus far for the long waves and short waves, complies with the Schrodinger. The problems may arise when we analyze ultrashort pulses rather then waves assumed to go to infinity, and we must analyze the higher order terms. The goal is to be able to extract both the nonlinear *and* the linear components even in a highly dispersive medium for the duration of the pulse, be it femto or micro. A full understanding and dissecting of the picosecond and femtosecond pulses will most likely require incorporating quantum physics into analytic solutions analogous to the N=2 case derived above.

Acknowledgement: The author wishes to thank Dr. M.J. Potasek for assistance in this effort.

REFERENCES

[1] **G.P. Agrawal**, (1995) Nonlinear Fiber Optics. Academic Press, Inc., San Diego.

[2] **R.S. Tasgal and M.J. Potasek** (1992*) Soliton solutions to coupled higher-order nonlinear Schrodinger equations*, J. Math. Phys., **33**, 1208-1217.

[3] **D. Marcuse** (1990) Theory of Dielectric Optical Waveguides. Academic Press, Inc., New York.

[4] **M.J. Ablowitz and H. Segur**, (1981) Solitons and the Inverse Scattering Transform. Siam, Philadelphia.

[5] **R.W. Boyd**, (1992) Nonlinear Optics. Academic Press, Inc., San Diego

[6] **A.E.Paul and M.J.Potasek** (1996) *Numerical Solution of the Nonlinear Schrodinger equation for Optical Pulse Propagation,* to be published.

[7] **M.J. Ablowitz and P.A. Clarkson** (1991) Solitons, Nonlinear Evolution Equations and Inverse Scattering. Cambridge University Press, Great Britain.

[8] **G.L. Lamb, Jr.** (1980) Elements of Soliton Theory. John Wiley & Sons, New York.

[9] **M.J. Potasek and M. Tabor** (1991) *Exact solutions for an extended nonlinear Schrodinger equation*, Phys. Lett. A, Vol 154, Num. **9**, 449-452.

[10] **A.C.Newell and J.V.Moloney** (1992) Nonlinear Optics. Addison-Wesley Publishing Company, California.

[11] **J. Satsuma and N. Yajima,** (1974) Prog. Theor. Phys. Suppl. 55, 284.

PRELIMINARY SPECIFICATIONS FOR SCREEN AND ANIMATION FOR INSTRUCTIONAL SIMULATION SOFTWARE DEMO

Jennifer L. Day
Department of Learning and Instructional Technology

Arizona State University
Tempe, Arizona 85287-0611

Final Report for:
Summer Graduate Student Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Armstrong Laboratory

August 1996

PRELIMINARY SPECIFICATIONS FOR SCREEN AND ANIMATIONS FOR INSTRUCTIONAL SIMULATION SOFTWARE DEMO

Jennifer L. Day

Introduction

The documents contained in this report are preliminary specifications in an effort to develop a proof of concept demo which will serve to demonstrate the capabilities and benefits of a desktop instructional simulation for Basic Fighter Maneuver (BFM) training. With issues of cost and safety as well as the call for more effective and efficient training, there is an increasing demand for more ground-based training and practice (Mattoon, 1995).

A discussion of the many benefits of desktop instructional simulation to pilot training are beyond the scope of this report. However, a few key benefits are mentioned here. In contrast to traditional hand gestures and planes on sticks to illustrate maneuvers, the use of three-dimensional animation, audio and simulation can provide students with the visual-specific information and situational contexts for the integration of cognitive and perceptual skills for high performance in dynamic flight environments (Andrews, D.H., Edwards, B.J., Mattoon, J.S., Thurman, R.A., Shinn, D.R., Carroll, L.A., 1995). In addition, the instructional simulation can allow both the student and the instructor pilot (IP) the means to configure individualized supplemental instruction and practice for use in and out of the classroom. This capability can be a factor in helping to reduce the IP's workload. Furthermore, complex flying tasks and concepts can be broken down into subtasks, which can reduce the difficulty students have in understanding and performing the tasks (Andrews, et al., 1996).

The specs for the demo prototype included within this document, focus primarily on the selected objectives of BFM, with an emphasis on Offensive Basic Fighter Maneuvers (OBFM). This work reflects the effort to date. Further development and modification of these and added documents is expected for optimal integration of additional aspects of BFM instruction. Evaluation of the demo will be reported at a later date.

Rationale for BFM Instructional Simulation Software

Problem

Student pilots in the fighter pilot career track are expected to perform complex situational flying tasks as outlined in their BFM instructional courses. However, they experience difficulty in understanding and visualizing the dynamic and rapidly changing nature of these concepts which is necessary for optimal performance of these tasks.

Why Important

This problem deserves careful attention if there is to be an increase in the efficiency and effectiveness of BFM training and performance. In-air training time is increasingly limited by cost and maintenance of aircraft and cannot be optimally utilized if it is taken up with remedial instruction. With less in-air training time, yet the consistent requirement for BFM graduates to demonstrate a high proficiency in performing complex BFM tasks, there is the need for the development of alternative ground-based instructional technologies which will afford student pilots dynamic, interactive instruction and practice of BFM concepts and tasks.

Proposed Solution

Armstrong Laboratory intends to develop and produce a desktop modeling and simulation BFM training software which can be delivered on high-end Macintosh or PC-based laptop computers which will be available to student fighter pilots for individual use inside and outside of the classroom environment. This software can be configured by both instructor and student to supplement BFM instruction and can be interactive with appropriate instruction, practice and feedback to assist the user in attaining the program objectives.

Front-End Analysis

In preparation for the development of content and technical specifications for the BFM instructional simulation software, a front-end analysis was performed to determine pertinent learner capabilities, setting components and instructional learner objectives. A number of subject matter experts were contacted to provide relevant information on content, target audience, technical aspects. In addition, print and non-print materials were consulted to provide specific information on the BFM instructional content. A list of all resources can be seen in Appendix A.

Audience and Setting Analysis

Audience

This instruction is aimed at Specialized Undergraduate Pilot Training (SUPT) Air Force graduates which are selected for the fighter pilot career track. These students are: (a) Top of their class, well-educated and intellectually bright; (b) experienced in the use of basic computer skills; (c) self-motivated and capable of accessing information on their own; (d) very interested in and comfortable with technology's role in the context of pilot instruction; (e) more interested in techniques rather than processes; (f) intensely aversive to failure.

Setting

The BFM instructional simulation demo will be delivered individually on PC-based laptops provided for each student pilot and is envisioned as being used in three capacities: a) concept presentation, b) classroom, and c) practice/rehearsal. The settings for use will include

classroom, lab or home. In order to facilitate use of the software's audio capabilities within these different settings, the user will be provided with headphones.

Instructional Objectives

Overall Objectives for Instruction

The overall goal for this instruction is three-fold: (a) To save money through decreasing necessary teaching time, (b) to facilitate optimal understanding of BFM in minimal time, and (c) to facilitate a higher level of proficiency in BFM task performance.

Specific Behavioral Objectives

The specific behavioral objectives reflect the current requirements of student pilots within BFM instruction and training. Given a variety of verbal data (e.g. current airspeed, turn rate, etc.) and static and dynamic visual examples (2D and 3D animations or video clips of simulation training flights) that depict OBFM situations, the learner will: (a) Recognize and identify the best choice of immediate objectives (e.g. increase/decrease closure, range, aspect) for a given set of controllable OBFM situations (student controls speed of event progress)--what, when, where, and how to close on an opponent as a function of BFM dynamics; (b) recognize and identify the consequences of each maneuver decision or action that is performed (what will be gained/lost) a given set of controllable BFM situations; and (c) execute timely adjustments in strategy by making the correct choice among alternatives in real time (time contingency) and by manipulating three-dimensional aircraft models on screen (spatial contingency)--adjust inputs such as attitude, compass heading, and speed.

Media Selection Rationale

For the BFM instructional simulation demo, the Armstrong R & D lab has selected desktop computer-based instructional simulation developed on the Power PC Macintosh, using a possible combination of Director, Soft Image, and C programming as authoring systems/tools for delivery on a high-end Macintosh or PC-based laptop. This decision was based upon the characteristics of the instructional content, learner, and objectives which are briefly discussed below.

Content

Text, audio and graphics are heavily integrated as well as demanding a high degree of interactivity with the user, both of which require multimedia simulation capabilities for effective delivery of instruction and practice. While the content is laid out in a linear fashion, instructional software can allow the individual user the flexibility of accessing necessary segments in a non-linear sequence. Simulation capabilities within the media can provide the learner with dynamic interaction of concepts which are difficult to visualize. In addition, this media makes it possible for complex tasks to be broken down into part-task execution, allowing the user to move smoothly from no knowledge of a task to a basic understanding, and eventually, to high proficiency of one or more complex task components.

Audience

The goal of the instructional simulation is to provide the users with a context in which they can apply selected BFM skills and knowledge in a manner consistent with real life application. On individual personal computers, the users will be able to move through the instruction at their own rate of speed. Part of the audience may only need instruction or review of select segments of the program. This medium will allow them to adapt the program to meet their needs. A computer-based medium will provide the audience with a user-friendly, interactive interface for becoming familiar with the content and achieving program objectives.

Objectives

The objectives primarily call for selected responses (or selected parameters for responses). A computer-based instructional simulation medium is ideal for this type of practice activity as it can model real life application of targeted skills. The computer-based medium can provide suitable practice and feedback to insure the learner's attainment of the instructional objectives. Practice and feedback for objectives can be responsive and individualized to each learner in an computer-based instructional environment.

Treatment

This instructional software will be designed to instruct student fighter pilots in the identification, recognition and execution of targeted BFM concepts and tasks. A combination of three-dimensional animation, audio, text and simulation will be used to motivate students to succeed in acquiring the knowledge and skills necessary to effectively and efficiently perform specific BFM tasks. Further development of instructional content, practice, feedback and data collection is needed before additional specific description of a finished product (i.e. layout of instruction, practice and feedback, and evaluation) can be given. Efforts for this development are currently underway.

Results

Instructional Content

The content developed to date includes initial instruction for the user on how to use the software as well as an overview of instructional objectives, content and strategies. Subsequent instruction introduces general BFM principles of positional geometry, control zone, weapons parameters, turning room and turning circles. The final section of instruction deals with general principles of OBFM and an "ideal" OBFM training exercise. Three basic training aspects are covered: (a) What the offender should do in this given canned OBFM exercise (including the visual cues of aspect, speed and range), (b) the common errors made by beginning pilots in this type of OBFM exercise and (c) how to correct these errors. The content outline for the instruction developed to date is listed in Appendix B.

Specifications

Screen interfacing specifications describe the layout and function of navigational buttons and animation and text fields for the BFM instructional demo (see Figures 1 through 6 in Appendix C). In addition, technical specifications have been designed both in drawing and animation form to provide the animation developers a "blueprint" to use in creating the animations for the BFM instructional demo. Animation specifications are provided for four scenarios: (a) The canned OBFM training setup (see Figures 7 and 8 in Appendix C), (b) what the offender should do (see Figures 9 through 12 in Appendix C), (c) three common errors made in this type of training exercise (see Figures 13, 15 and 17 in Appendix C), and (d) how to fix those errors (see Figures 14, 16, and 18 in Appendix C). The common errors and their corrections are grouped together respectively for ease in instructional transitioning. It is expected that additional animation specifications will be developed and included (e.g. the control zone and weapons engagement parameters) for use in the instructional software demo .

Appendix A
Resource Bibliography

People

Debra Bolin
Presentation Graphics Specialist
Hughes Training, Inc.
Mesa, Arizona 85206-0904

Dr. Rebecca Brooks
Research Psychologist, Flight Simulation and Training
Aircrew Training Research Division
Armstrong Laboratory
Mesa, Arizona 85206-0904

Dr. Bernell Edwards
Research Psychologist
Aircrew Training Research Division
Armstrong Laboratory
Mesa, Arizona 85206-0904

Scott Gagel
Videographer
Hughes Training, Inc.
Mesa, Arizona 85206-0904

Justine Good, Captain, USAF
Modeling and Simulation Development Pilot
Engineering Development Division
Armstrong Laboratory
Mesa, Arizona 85206-0904

Margie McConnon
Computer Graphics Designer
Hughes Training, Inc.
Mesa, Arizona 85206-0904

Dr. Joseph Mattoon
Personnel Research Psychologist
Aircrew Training Research Division
Armstrong Laboratory
Mesa, Arizona 85206-0904

Dr. Byron Pierce
Research Psychologist, Visualization
Aircrew Training Research Division
Armstrong Laboratory
Mesa, Arizona 85206-0904

William B. Raspotnik
Training Analyst, Pilot
Hughes Training, Inc.
Mesa, Arizona 85206-0904

Reid Reasor, Captain, USAF
Modeling and Simulation Development Pilot
Engineering Development Division
Armstrong Laboratory
Mesa, Arizona 85206-0904

Robert Robbins
Training Analyst, Pilot
Hughes Training, Inc.
Mesa, Arizona 85206-0904

Trish Russo
Research Scientist, Virtual Reality
Hughes Training, Inc.
Mesa, Arizona 85206-0904

Robin Smith
Instructional Designer, Digital Imaging and Visualization
Department of Curriculum and Instruction
Arizona State University
Tempe, Arizona 85287-0111

Dr. Richard Thurman
Research Psychologist, Training and Simulation
Aircrew Training Research Division
Armstrong Laboratory
Mesa, Arizona 85206-0904

Brian Watkins
Imaging/Multimedia Specialist
Hughes Training, Inc.
Mesa, Arizona 85206-0904

Print Resources

Department of the Air Force. (1996)
19 AF Instructor Guide B/F-V5A-K-AA-IG
Washington, D.C.: Headquarters US Air Force.

Department of the Air Force. (1995)
AF Manual 11-238, Volume 1
Washington, D.C.: Headquarters US Air Force.

Department of the Air Force. (1995)
AF Manual 11-238, Volume 2
Washington, D.C.: Headquarters US Air Force

Department of the Air Force. (1993)
ENJJPT Study Guide/Workbook P-V4A-N3-FM0101
80th Flying Training Wing
Sheppard AFB, Texas 76311-6424

Department of the Air Force. (1993)
AETC Instructor Guide B/F-V5A-K-AA-IG
Washington, D.C.: Headquarters US Air Force

Department of the Air Force. (1993)
AETC Workbook B/F-V5A-K-AA-WB
Washington, D.C.: Headquarters US Air Force

Department of the Air Force. (1992)
AF Manual 3-3.
Washington, D.C.: Headquarters US Air Force

Non Print Resources

Mattoon, J. S. (Producer). (1996).
The Electronic Classroom
[Video]
Armstrong Laboratory
Mesa, Arizona 85206-0904

Reasor, R. W. (1996).
BFM Visualizations: Animations of BFM Concepts
[computer software: .AVI files]
Armstrong Laboratory
Mesa, Arizona 85206-0904

Reasor, R.W. (Instructor), & Gagel, S. (Director/Producer) (1996).
Civilian Fighter School: Intro to BFM
[Video]
Armstrong Laboratory
Mesa, Arizona 85206-0904

Appendix B
Content Outline for Demo

1. Introduction (the first time, second time user)
    - 1.1 Computer operation assistance
    - 1.2 Assistance with the navigation buttons
    - 1.3 Introduction to the program objectives
    - 1.4 Introduction to the program content, strategies
2. General principles of BFM
    - 2.1 Introduction
    - 2.2 Positional geometry
        - 2.2.1 Description of angular relationships and positional advantage
            - 2.2.1.1 Range
            - 2.2.1.2 Aspect angle
            - 2.2.1.3 Angle-off
        - 2.2.2 Description of attack pursuit courses
            - 2.2.2.1 Lead
            - 2.2.2.2 Lag
            - 2.2.2.3 Pure
    - 2.3 Control zone
    - 2.4 Weapons parameters
        - 2.4.1 Guns
            - 2.4.1.1 minimum range
            - 2.4.1.2 maximum range
        - 2.4.2 Aim 9
            - 2.4.2.1 minimum range
            - 2.4.2.2 maximum range
        - 2.4.3 Aim 7
            - 2.4.3.1 minimum range
            - 2.4.3.2 maximum range

    - 2.5 Turning room and turning circles
        - 2.5.1 Rate
        - 2.5.2 Radius
        - 2.5.3 Corner velocity (energy versus nose position)
        - 2.5.4 Lateral and vertical turning room

3. Set-up for transition to turn circle exercise
    3.1 "Ideal" OBFM training exercise
        3.1.1 Visual Cues
            3.1.1.1 Aspect
            3.1.1.2 Range
            3.1.1.3 Closure
    3.2 Common errors made by offenders in canned OBFM training scenarios and how to correct them
        3.2.1 Error at point A: pure pursuit
            3.2.1.1 Visual cues
                3.2.1.1.1 Aspect
                3.2.1.1.2 Range
                3.2.1.1.3 Closure
        3.2.2 How to correct pure pursuit error at point A
            3.2.2.1 Visual cues
                3.2.2.1.1 Aspect
                3.2.2.1.2 Range
                3.2.2.1.3 Closure
        3.2.3 Error at point B: turning too early
            3.2.3.1 Visual cues
                3.2.3.1.1 Aspect
                3.2.3.1.2 Range
                3.2.3.1.3 Closure
        3.2.4 How to correct turning too early at point B
            3.2.4.1 Visual cues
                3.2.4.1.1 Aspect
                3.2.4.1.2 Range
                3.2.4.1.3 Closure
        3.2.5 Error at point B: turning too late
            3.2.5.1 Visual cues
                3.2.5.1.1 Aspect
                3.2.5.1.2 Range
                3.2.5.1.3 Closure
        3.2.6 How to correct turning too late at point B
            3.2.6.1 Visual cues
                3.2.6.1.1 Aspect
                3.2.6.1.2 Range
                3.2.6.1.3 Closure

Appendix C

Screen Specifications and Sample Screens

Note. In reading the subsequent animation specifications and notes, the offender is sometimes referred to as the attacker and the defender is sometimes referred to as the bandit. For ease in differentiating between them, the offender is represented in white and the defender in black.)

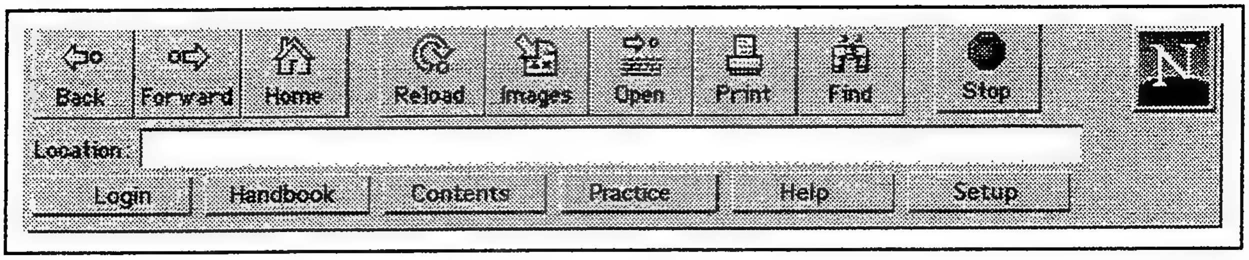


Figure 1. Netscape Navigational Button Bar

Button Functions:

Back: Takes user to previous screen.

Forward: Takes user to next screen.

Home: Takes user to Main Menu.

Reload: Reloads current screen to original status of present instructional section.

Images: (Function not determined at this time.)

Open: Opens pop-up text box for user to type in the name or section number of the desired instructional module or section (as opposed to going back to Main Menu). Upon clicking return, user is taken to the specific section screen.

Print: Performs screen capture of current screen and prints.

Find: Functions like a search engine. A pop up text box appears in which user types in a key word or phrase he wishes to locate within the current instructional section. Upon clicking on a "find it" button, the user is taken to a list of hypertext links with short descriptors. User clicks on desired link which takes him to specific screen where key word or phrase is used.

Stop: Deactivates animation, audio and/or current transition.

Location: Displays the section or unit where user currently is within the instructional module.

Login: A toggle button (login/logout); login--user is presented with two options: 1) New user--a pop-up text field at beginning of session for user to enter name, i.d. number and any other information useful for data collection purposes. 2) Repeat user--a pop-up text field in which user enters his password (repeat users can then go back to main menu or to resume where they left off.

Handbook: Takes user to an indexed navigational and technical guide for the program, complete with troubleshooting and FAQ (frequently asked questions).

Contents: Takes user to a hypertext linked index for instructional content of program.

Practice: Takes user to set (default) practice sections relative based on the section or unit he is working on or has just completed. Or the instructor/student can configure level of mastery, specific configurations for types of practice items (true/false, drag and drop, multiple choice, simulation, etc.) as well as specific objectives to be included or omitted in practice items.

Help: Opens full screen window with a topics list and alphabet linked to key terms and concepts (hypertext links to audio, schematic, animated illustrations)

Setup: Displays a pop-up menu with animation settings (audio: on/off, WEZ: on/off, pilot: offender/defender, schematics: on/off).

offender/defender, schematics: on/off).

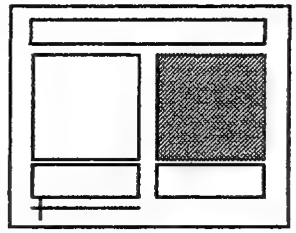


Figure 2. Animation Screen: Outside View Animation Field

Location:

The Outside View field will be on the right of the animation screen, opposite the Pilot View field. (See small picture at top right for screen representation.) The animation of the Outside View will run synchronous in real time with that of the Pilot View (which will be either the attacker or the defender perspective).

General Description:

Narration will introduce each training context and the Outside View animation will open and start concurrently with Attacker view in real time. Animation will begin with two F-16 aircraft setting up for offensive basic fighter maneuver (OBFM) training situation. The user will see the depicted situation from an overhead (God's-eye) view. Both aircraft will move around the defender's turn circle, which will be enlarged from actual scale for visibility. The aircraft must be visible with enough detail to show pitch, yaw and roll. Altitude of aircraft will remain constant at 18,000' above ground level. The beginning frame of each scenario will start with the attacker and bandit setting up for the OBFM training maneuver. (Set up details to follow.) Animation will proceed with a variety of OBFM training situations. The focus of these will be to depict what the offender's visual cues will be, followed by what he should maneuver and finally the common errors made and how to correct them. Ending frame will show attacker firing at the bandit with sound effects and explosion or firing with sound effects and screen dissolve into introduction of subsequent animation scenario.

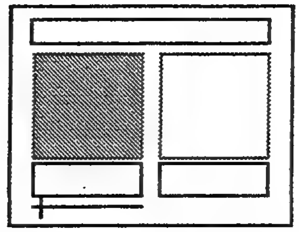


Figure 3. Animation Screen: Pilot View Animation Field

Location:
The Pilot View will be adjacent to the Outside View field, on the left side of the animation screen. This field will show animations of both the Offender View and Defender View at different times.

General Description:
The animations in the Pilot View field will be the same animation as the Outside View animation, but from different perspectives. Accuracy of the offender and defender aircraft detail, maneuvering and timing within each scenario is crucial to these animations. User must be able to perceive that the distance between the aircraft, speed, pitch, yaw and roll of both aircraft are real. Certain heads up display (HUD) data will be displayed numerically at the bottom of both pilot views. For the offender: aspect, range and closure; and for the defender: aspect, range, heading crossing angle, closure and line of sight (LOS).

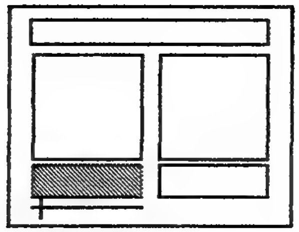


Figure 4. Animation Screen: Schematic View Field

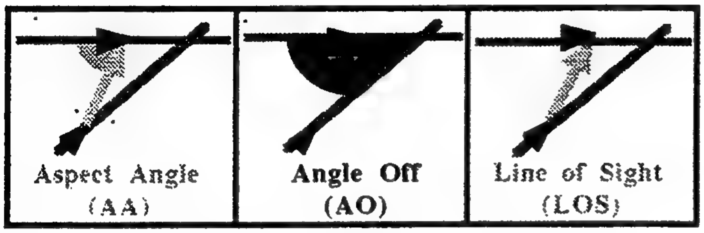


Figure 5. Schematic View Field Example

Location:
Beneath Pilot View field.

Description:
The animation of the schematics will be created as part of the same animation as Outside and Pilot View animations. User will be instructed to select the configuration of the animations on the screen. One option will be turning on the schematic animation by clicking on the schematic icon in the lower left corner of the Schematic field. Once the schematic icon has been clicked and the start button has been selected, three adjacent line animations will appear and will run concurrently in real-time correspondence with the Outside and Pilot View animations. Angles will increase and decrease and LOS line will lengthen or shorten in real-time correspondence to Outside and Pilot View animations.

Color and Size:
Colors of lines, angles aircraft and text must be clear and contrasted for ease of reading. Each of the three schematics will be surrounded by a dark-colored frame. Background color of schematic animation will be white. Longitudinal axis lines of both aircraft will be the same color (suggested green). Angles will be shaded with different transparent color. LOS line will be a contrasting solid color. Lines should be 2 point. or 3 point. in width.

Text:
Text will be colored to correspond to the angle (different for AA and AO) and to the line for Line of Sight. Any instructional text will be in 12 point, Palantino or Times font (titles in 14 or 18 bold).

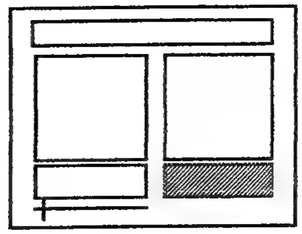
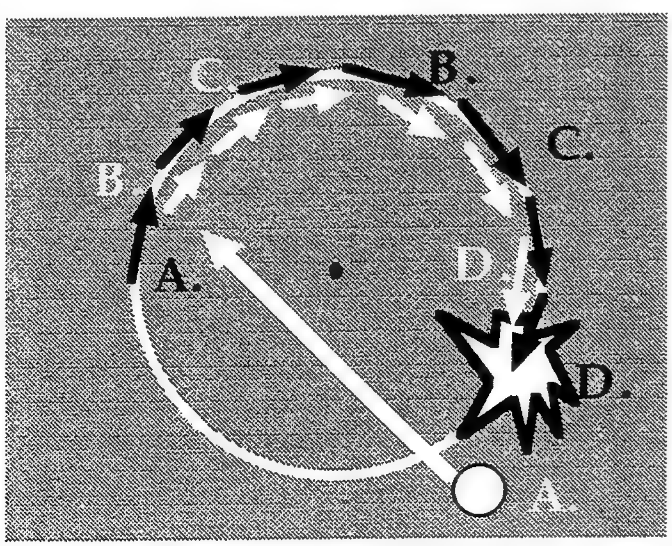


Figure 6. Animation Screen: Text Field

Location:
Beneath Outside View Field.

Description:
Text field will a scrollable field of hypertext terms or concepts specified by the designer. Once user clicks on text field (scrollbar or link) all animations stop. Upon clicking a hypertext link, a pop-up box will appear with the term's definition or explanation. Clicking on the pop-up box will deactivate it and resume animation.

Font, size, color:
Font used will be Palantino or Times or Times. Size of font will be 12-point and color used for text will be black for non-hypertext links. Active hypertext links will be underlined and displayed in blue and after selected, will be displayed in red.



The animation for the first Outside View should depict the ideal OBFM training scenario in which the offender ends with the following:

Aspect:	<20° to 25°
Closure:	10% of his airspeed
Range:	1,500' to 2,000'

Figure 7. Ideal OBFM Scenario

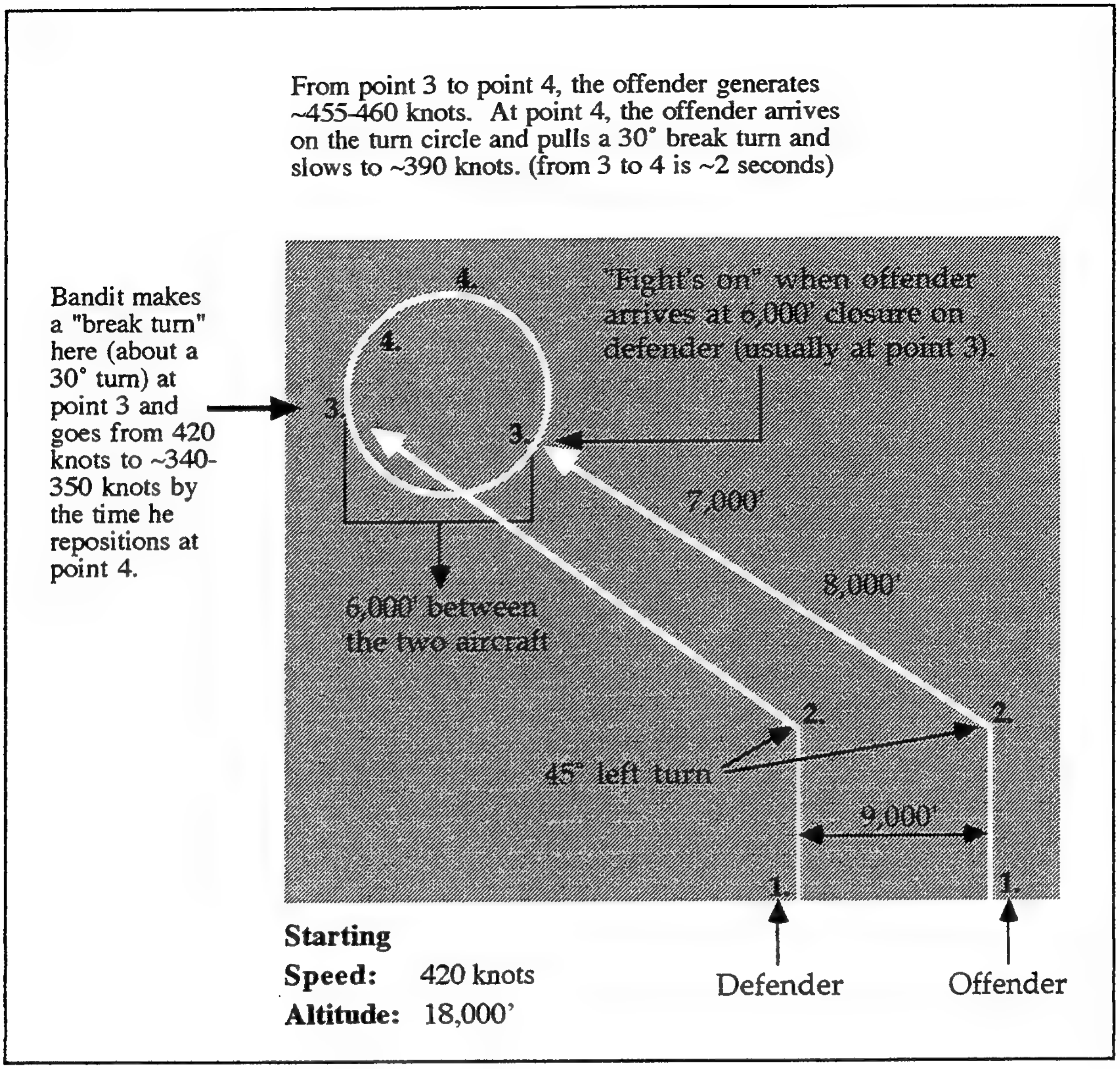
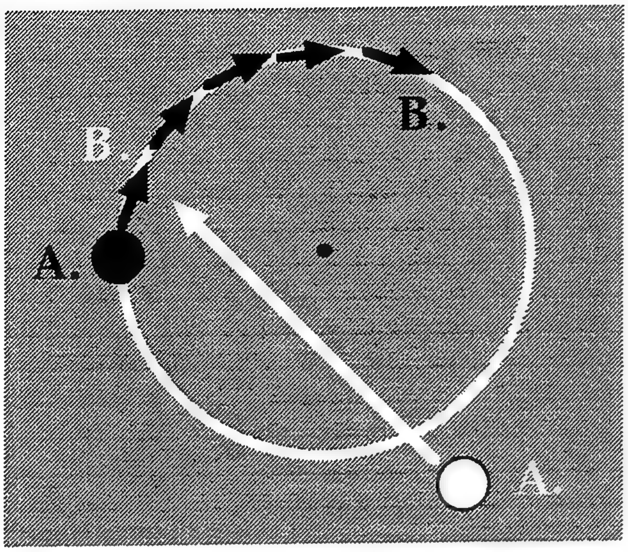


Figure 8. "Ideal" OBFM Training Setup and Exercise



The defender initiates his break turn of 30° aspect at point A, slowing him to 340 to 350 knots by point B. The offender gains to 455 to 460 knots from point A to point B. At their respective endpoints, range is ~4200'. (Note: in real-time, this transition lasts 2 seconds.)

Figure 9. Visual Cues for Second Animation Sequence

The offender will see the bandit moving gradually across the offender's canopy. Then there will be a sudden increase in the bandit's speed across the offender's canopy. At that moment, the offender has arrived on the bandit's turn circle and must pull a 30° turn to the right.

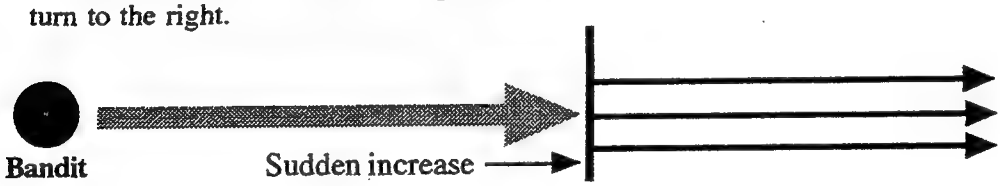
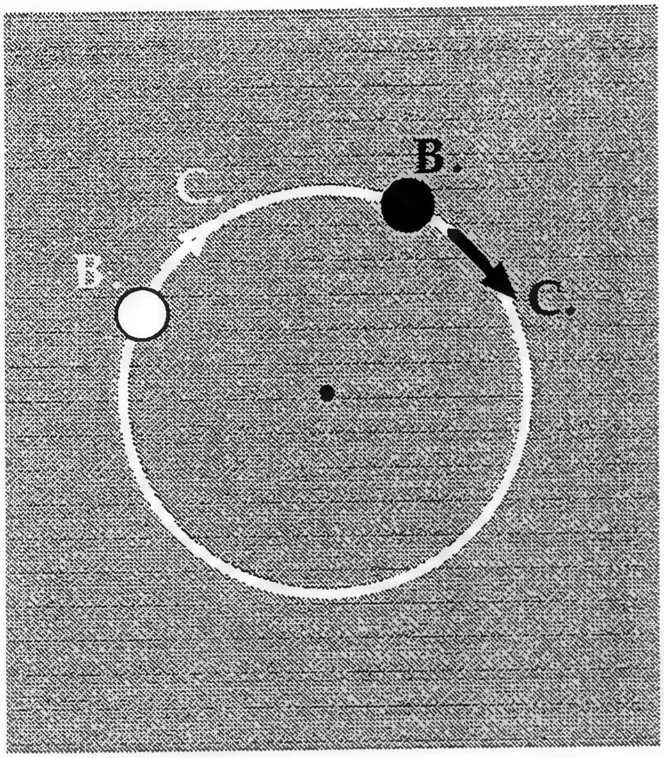
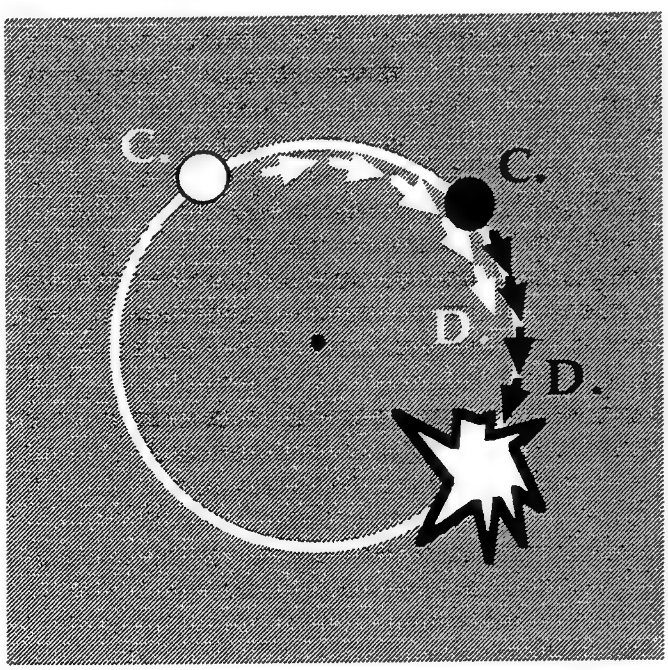


Figure 10. Break in Line of Sight



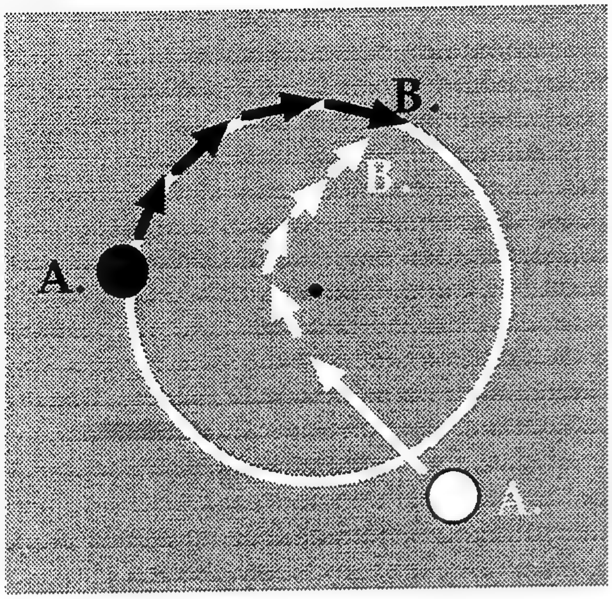
The offender pulls a 30° break turn at point B which slows him to 390 knots by point C. Animation from offender's view needs to illustrate the "break" or sudden increase in the bandit's LOS (left to right movement of the bandit's aircraft across the offender's canopy). That is the visual cue the offender watches for to tell him when to break turn. The defender continues around the turn circle from point B to point C at 340 - 350 knots, with a 30° angle on the offender.
Distance between the aircraft at their respective endpoints is ~3600'.

Figure 11. Aspect, speed and range from points B to C.



The offender maintains the energy advantage from point C to point D with speeds of 390 to 400 knots (which means he has a minimum of +50 knots closure on the defender).
The defender slows to 340 knots by point D. As both aircraft move toward their respective endpoints, aspect decreases and is 20°- 25°. Range between aircraft has decreased to 1500'. At the end of this scenario, the offender is shown shooting the guns at the defender and then separates by rolling out.

Figure 12. Aspect, Speed and Range from points B to C.



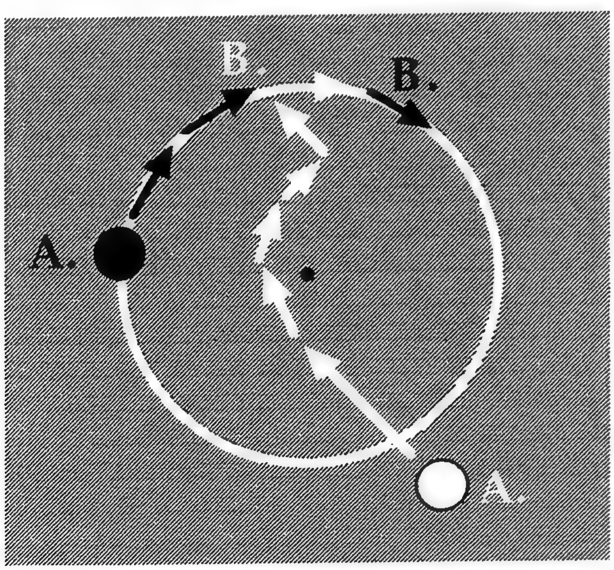
In this scenario, the offender arrives at point A and flies with his nose pointed directly at the bandit (pure pursuit) throughout the defender's turn. By the time the offender meets the bandit, the following conditions should be evident (pertaining to the offender) in the animations from all perspectives:

Aspect: very high, ~ 90°

Speed: closure of ~400 knots

Range: rapidly decreasing

Figure 13. Offensive Error at Point A: Turning Too Early Prior to LOS Break



Animation for this sequence must illustrate the offender turning back out and repositioning on the turn circle. The end of this sequence should show the following specs for the offender:

Aspect: 20° - 25° behind bandit

Speed: 390 - 400 knots

Range: 1,500' - 2,000'

Figure 14. How to Correct Pure Pursuit Error

PROJECTED IMPACT OF A PROTOCOL ADJUSTMENT ON THE INVALID OUTCOME RATE OF THE USAF CYCLE ERGOMETRY ASSESSMENT

Gerald DeWolfe
Research Associate
Department of Kinesiology and Health Education

Franklin Flatten
Research Associate
Department of Kinesiology and Health Education

Stefan Constable, Ph.D.
Exercise Physiologist
Armstrong Laboratories

Lt. Ryan O'Dowd
Air Force Fitness Program

Lt. Dean Richardson
Air Force Fitness Program

The University of Texas at Austin
Department of Kinesiology and Health Education
222 Bellmont Hall
Austin, TX 78712

HSC/YAM
Air Force Fitness Program
8107 13 St.
Brooks AFB, TX 78235-5218

Final Report for:
Graduate Student Research Program
Armstrong Laboratory

Sponsored By:
Air Force Office of Scientific Research
Bolling Air Force Base, DC
and
Armstrong Laboratories

October 1996

PROJECTED IMPACT OF A PROTOCOL ADJUSTMENT ON THE INVALID OUTCOME RATE OF THE USAF CYCLE ERGOMETRY ASSESSMENT

Gerald DeWolfe, Research Assistant, Department of Kinesiology and Health Education, The University of Texas at Austin

Abstract

Pass, Fail and Invalid outcomes of the US Air Force's Cycle Ergometry Assessment were analyzed from data collected at five AF bases. An Invalid results when the heart rate (HR) response falls outside of the parameters set forth for the assessment (i.e. HR too high or HR below 125 beats per minute (bpm)), the subject requests termination of the assessment, or an error occurs due to either equipment failure or assessment administrator error. Of all tests analyzed 16.4% tests (1548 of 9437) resulted in an Invalid outcome (74.0% Passed, 9.6% Failed). The total Invalid outcomes were then sorted by (seven) categories, and excessive heart rate (HR) i.e., Category 1, accounted for the greatest percentage of Invalids (39.7%). These Invalids are primarily due to the projected workload (WL) being too high. Most subjects who re-test at a lower WL setting receive a score. Therefore, lowering the HR range required for an increase in the WL during the assessment should maintain the HR below the 85% HRmax cutoff and allow for a score to be assessed. Further in depth analysis suggested that a 10-bpm adjustment (decrease) in minutes 3 and 4 of the WL adjustment criteria would potentially reduce the Invalid rate by, at best, only 1.6% of total tests (14.8% total Invalids). This estimate was formulated because most subjects who receive the Category 1 Invalid do not receive any WL progression in minutes 3, 4, or 5. Therefore, no adjustment to the required HR response would affect the WL. The total Invalid rate may be further decreased by other testing protocol adjustments.

INTRODUCTION

The need to accurately assess the fitness level of the Air Force (AF) population has been addressed with a submaximal cycle ergometry (CE) assessment. For a submaximal assessment to predict maximal oxygen consumption ($VO_{2\,max}$) there must be an interval period in which the HR is assessed at steady state. The HR range for the AF CE assessment during this interval is a minimum of 125 bpm, to a maximum of 85% of HR maximum (HRm); i.e. 85% of HRm calculated as [(220 - age) X .85]. If an individuals HR response falls outside of this range, $VO_{2\,max}$ may not be as accurately predicted. The possible outcomes of the AF fitness assessment are Pass, Fail, or Invalid. When an individual's HR response falls outside of the designated range, the assessment is categorized as an Invalid. At present, the CE assessment too often results in an Invalid assessment, specifically Category 1 or high HR, and no "score" is assessed. The subject must then be re-tested on a subsequent day.

Anecdotal evidence from fitness assessment personnel first suggested a majority of Invalid assessments were due to subjects exceeding 85% HRm (Category 1 Invalid). Excessive Invalid assessments and the resulting need for a re-assessment present an unwanted drain on manpower and resources, as well as morale. It was therefore postulated that subjects who received an Invalid Category 1 outcome may have the greatest potential to instead receive a score Pass or Fail) after an adjustment to the workload progression portion of the CE 7-4

protocol is made. Therefore, the purpose of this study was twofold: 1) to determine the rates of Pass, Fail, and especially Invalid assessment outcomes across Categories 1-7, and 2) to analyze the potential impact of a 10 bpm reduction of the heart rate parameters, which determine the workload progression portion of the evaluation, during minutes 3 and 4 on the final assessment outcome (i.e. a Pass, Fail, or Invalid result).

A follow-up study will compare the current assessment protocol to two proposed protocols in an attempt to reduce the overall number of Invalid assessments. The two protocols have been

designated Protocol A and B. Protocol A will alter the computer logic to make it more difficult for a subject to receive a 1.0 kilopond (Kp) or 0.5 Kp workload progression, i.e. lower the minimum HR needed to receive a workload increase (Appendix 1, Part B). Protocol B will lengthen each of the three stages at which workload progression occurs by 1 minute , thereby allowing more time to achieve a steady state HR. Only the potential impact of Protocol A will be discussed further in this report.

METHODS

Available information on fiscal year 1996AF submaximal CE assessments from five bases was collected and analyzed. For this initial evaluation, data is based on the number of total assessments with special interest in service members who received Category 1 Invalid outcomes. These numbers include the same individuals who took repeat assessments. All results are calculated from the combined male and female data unless otherwise noted.

Total assessments (all Pass, Fail, Invalid and re-tests) from Brooks AFB (n=530), Kelly AFB (n=2118), Lackland AFB(n=2191), Patrick AFB (n=2363), and Randolph AFB (n=2235) were sorted and the Pass, Fail, and Invalid rates were determined. Invalid frequencies for the seven Invalid Categories were also calculated. The seven Categories were deliniated by the following:

1) HR exceeds 85% of maximum (HRm; based on 220-age);
2) HR does not reach 125 beats per minute (bpm) in the last minute of the assessment;
3) HR varied more than 3 bpm in the final 2 minutes;
4) Subject could not maintain 50 revolutions per minute (rpm);
5) Rating of Perceived Exertion (RPE) exceeds 15;
6) Subject requested termination of the assessment;
7) Other.

Category 5 (RPE exceeds 15) was deleted in April of 1996.

Combined Category 1 Invalid assessment data from Brooks, Kelly, and Randolph AFB were compiled and further analyzed by HR response and WL progression during minutes 3, 4, and 5 (Tables 2-5). Individual Invalid data from Brooks, Kelly, and Randolph AFB is provided in Appendix 4-7. Due to the large time investment necessary to analyze the data in this manner, this analyses was not done for Lackland or Patrick AFB. However, assessment records for selected individuals from 12 other AF bases who had three Invalid outcomes were also analyzed. These data were separated by Invalid category and only the Category 1 Invalid

assessments were used for separate analyses. These service members WL progressions and HR responses during minutes 3, 4, and 5 were also determined (Table 6).

Re-test data from the five bases were also examined. For this analysis, assessments were separated by subject number so that the total number of subjects could be differentiated from the total number of assessments.

Subject data were downloaded from the FitSoft 2.0 database at four of the bases; while the fifth base, Patrick, uses AF 2000 software (Microfit, Inc.). Protocols and algorithms for Fitsoft 2.0 and AF 2000 are the same regardless of the software. All data were transferred to and sorted on Microsoft Access. Further analysis was performed with Microsoft Excel 5.0.

RESULTS

Invalid assessment outcomes from the five bases accounted for ≈16.4% of all assessments (n=1548 of 9437), while the percentage receiving a Pass was ≈74.0%(n=6985 of 9437), and ≈9.6% Failed (n=904 of 9437; Table 1). The Invalid assessment breakdown as a function of the total number of assessments evaluated is as follows (see Methods for category descriptions): Category 1 accounted for 6.5% of all assessments, Category 2 accounted for 1.3%, Category 3 accounted for 2.5%, Category 4 accounted for 0.08% ,Category 5 accounted for 1.5% (category deleted as of April 1996), Category 6 accounted for 0.3%, and Category 7 accounted for 4.3% of Invalid assessments (Table 1). Again, repeat test outcomes were not discriminated here.

Analysis of only initial assessment outcomes from Brooks, Kelly, and Randolph AFB were completed to determine if the analysis of total combined assessment data was a reasonable approxiamation of what occured on the initial evaluation. Invalid assessment outcomes from the three bases accounted for 16.2% of all assessments (n=660 of 4070), while the percentage receiving a Pass was 75.8%(n=3086 of 4070), and 8.0% Failed (n=324 of 4070; Table 1A). The Invalid assessment breakdown as a function of the total number of assessments evaluated is as follows (see Methods for category descriptions): Category 1 accounted for 6.4% of all assessments, Category 2 accounted for 0.5%, Category 3 accounted for 1.9%, Category 4 accounted for 0.05% ,Category 5 accounted for 2.3% (category deleted as of April 1996), Category 6 accounted for 0.2%, and Category 7 accounted for 4.9% of Invalid assessments (Table 1A).

Combined Category 1 data from Brooks, Kelly, and Randolph AFB were examined by minute of assessment (n=279 for minute 3, n= 264 for minute 4, and n=255 for minute 5) and workload progression (Tables 2, 3 and 4). The number of assessments in each minute declines due to the early termination of some assessments generally because of subjects' HR being greater than 85% of predicted maximum. Category 1 Invalid assessments were reviewed in detail at these bases because the data suggests that changes to the current protocol which impact this category should reduce the greatest number of Invalid assessments (Figure 1). Table

2 shows that 49.8% of these assessments did not receive a workload increase in minute 3. The frequency of not receiving any WL progression increases dramatically in minutes 4 and 5 (81.1% and 72.5%, respectively; Tables 3 and 4). Overall, at roughly 67.4% of the 798 decision points (points during assessment when a WL progression could occur, i.e., minutes 3, 4, and 5) no WL increase was indicated.

The invalid assessments from Brooks, Kelly, and Randolph AFB (n=279) were further categorized by the magnitude of the HR response relative to the WL increases during minutes 3, 4, and 5 (Table 5). This breakdown was completed in order to estimate the potential impact of making the WL progression criteria more conservative, i.e. lowering the HR range to make it more difficult to receive a WL progression. Due to the excessive time needed for the analyses, it was not determined if those who received a WL progression in minute 3 also received a WL progression in minute 4 and/or minute 5 or vice versa. Results reported here are based on total assessment data and not on individual responses (the subject is counted as many times as they were re-assessed).

A smaller data base using individuals with three or more Invalid assessments was also used to evaluate the HR response to minutes 3, 4 and 5 of the assessment. The records for 46 subjects were evaluated and it was determined that of 138 assessments, 59 were identified as Category 1 Invalid (Table 6). It was not possible to distinguish between the annual assessment, first re-test, or the second re-test for this data. The data show that 76.3%, 94.8%, and 86.0% of these assessments had no workload progression at minute 3, 4, and 5, respectively. Of the original 59 Invalid assessments, one assessment was terminated before minute 4, and eight were terminated before minute 5. These numbers correspond to a total of 167 decision points. In 143 of these cases (85.6%) no workload progression was received. Any AF member receiving an Invalid must re-take the assessment. Data from the five bases revealed that of subjects who receive a Category 1, 2, 3, 4 or 6 Invalid on their first assessment, 55.8% (n=280) Pass on their first re-test, 22.1% (n=111) Fail, and only 22.1% (n=111) have a second Invalid result (Table 7). Category 5

and 7 Invalid assessments were excluded from the analysis because Category 5 was deleted as an option in April of 1996 and Category 7 is not indicative of subject response, but rather is due to equipment or Fitness Assessment Monitor (FAM) error. Thus, to more accurately evaluate the potential impact of Protocol A, only Invalid categories which are directly caused by or related to the protocol were included in the re-test analysis. Of the 111 individuals with an Invalid outcome on their first re-test, 70 had completed their second re-test with the following results: 44.3% (n=31) Pass, 20.0% (n=14) Fail, and 35.7% (n=25) had a third Invalid assessment. These numbers are only for re-tests after an Invalid and do not include re-tests after a Fail on the first re-test.

DISCUSSION

This study was primarily undertaken because of the perceived high incidence of Invalid fitness assessments due to HR above the accepted range (>85% HRm; Invalid Category 1). Our analysis of available data has shown that Category 1 Invalid assessments account for only 6.5% of all assessments at the five bases studied (n=9437; Table 1). Category 1 Invalid assessment outcomes (n=615) account for 39.7% of all Invalid assessments (Table 1, Figure 1). In other words, even though the percentage of total assessments accounted for by a Category 1 Invalid is lower than expected, the percentage of Invalid Category 1 assessments is still considerable. The impact of a modified protocol on reducing Invalid outcomes will therefore be lower than desired Even so, Protocol A may affect the largest single group of Invalids and therefore have a substantial impact on reducing the total number of Invalid assessments. This protocol change could possibly have some impact on Categories 2-4 and 6 as well. It is speculated that lowering the HR range needed to elicit an increase in WL may increase Category 2 Invalids, but may decrease the number of Category 3, 4 and 6 Invalids.

Protocol A is designed to affect the workload progression by making it more difficult for a subject to receive an increase in workload. For example: a 33 year old subject having a HR of 102 bpm at minute 3 in the current AF protocol would receive a 1 Kp progression, whereas in Protocol A the individual would receive a .5 Kp workload progression (see Appendix 2 for HR criteria), thereby keeping the HR lower. It is estimated that Protocol A could reduce the number of Category 1 Invalids outcomes by only 55.5% at the very best (38.1% of assessments possibly affected in minute 3 + 17.4% of assessments possibly affected in minute 4; Table 5). Therefore, Protocol A could reduce the total number of Category 1 Invalid assessments from 39.7% to 22.7% (From Tables 1 and 5: [(615-341)/(1548-341)](100)=22.7%). A reduction in Category 1 Invalid assessments from 39.7% to 22.7% could reduce the percentage of total Invalid assessments from 16.4% to 12.8%, thus potentially lowering the total number of Invalidassessments by 341 assessments, or 3.6%

It is important to note that multiple asssments (re-tests) by the same subject could not be discerned. Consequently, the predictions presented here are based on the *total* number of assessments. That is, without correction for the possible re-testing of subjects. Therefore, an individual receiving a workload progression in both minute 3 and 4 is evaluated as two assessments. This could easily lead to overestimation of the impact of Protocol A on the Invalid rate(s).

Evaluation of the initial assessment data (n=4070) from Brooks, Kelly, and Randolph AFB, excluding Categories 5 and 7, indicated that 90.2% of these assessments receive a score, 9.8% have an invalid outcome (Table 1A). This percentage was determined by subtracting Category 5 and 7 assessments (n=291) from the total number of Invalid assessments and the number of total assessments (i.e., (660-291) and (4070-291), respectively). 77.9% of individuals with an initial invalid assessment, excluding category 5 and 7, received a score on the first re-test (see Table 7). Therefore, 97.8% of all subjects receive a score within the first two assessments.

Analysis of 279 Category 1 Invalid assessments from Brooks AFB, Kelly AFB and Randolph AFB showed that only 50.2%, 18.9%, and 27.5% (Tables 2, 3, and 4, respectively) of assessments had a workload progression at minutes 3, 4, and 5, respectively. In comparison, the data for subjects with three Invalid assessments (Table 6) show that only 23.7%, 5.2%, and 14.0% of assessments have a workload progression at minutes 3, 4, and/or 5, respectively. This indicates a majority of subjects who receive an Invalid score are riding at or near the initial workload for the entire assessment (see Appendix 2). The initial workload is based on gender, age, weight, and self-reported activity level. While it is possible some individuals could receive a Passing score without an increase in workload, the score would probably indicate that they are in the lowest range of passing scores. As is shown in Table 7, 55.8% of first re-tests result in a Pass while 22.1% Fail. This would indicate that while a preponderance of those who receive an initial Invalid outcome can pass the assessment, it is generally only after the software initiates a lower WL allowing the heart rate to stay lower that they are able to pass (i.e., they are more

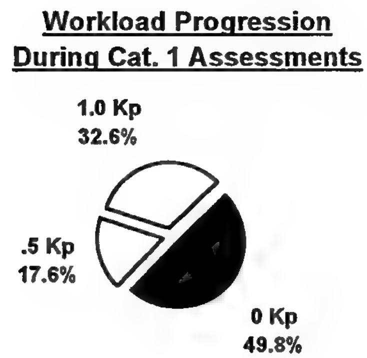
unfit since it takes a lower WL to keep their HR below the upper limit). Since the Fail rate is twice as high in this re-test group compared to the initial assessment outcomes, it appears that the first Invalid outcome is often masking what should be categorized as a Fail. This is an important consideration with regard to further protocol adjustments.

The second largest category of Invalids was Category 7 ("Other"). Data from Brooks, Kelly, Randolph, Lackland, and Patrick AFB's demonstrate that Category 7 Invalids make up 26.5% of all Invalid assessments, and 4.3% of total assessments (Table 1). Category 7 normally indicates FAM error and in very few cases equipment error. Most of the software problems, specific to the assessment, have been identified and corrected with the newest version of FitSoft (FitSoft 2.0), yet computer and equipment failures will continue to happen intermittently. HR monitors may "fail" when the battery runs low, when the monitor is not properly placed during subject preparation, or when the transmitter is too far away from the watch during the assessment. Tester error can include improper HR monitor operation or placement, as well as inaccurate data entry or work-load setting. More thorough training in CE and familiarity and knowledge of the typical responses to exercise may reduce the incidence of Category 7 Invalid outcomes by the FAM. Other factors such as scale calibration (body weight), higher or lower rpm, talking while cycling, self-reported activity level, fan availability, and room temperature all have an undetermined but possible impact on the assessment and may contribute to the high Invalid rate. However, other modest protocol adjustments, i.e. HR variability criteria and computer logic, minimum passing WL criteria and computer logic, etc, may offer the most fruitful and pramatic approach to further reduce the rate of repeated Invalid assessment outcomes.

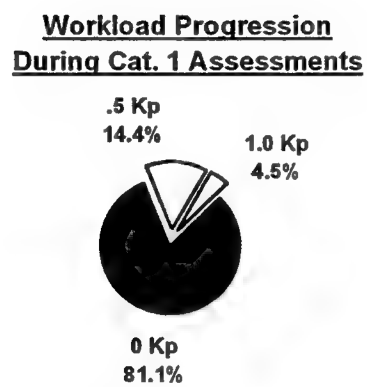
FIGURES

Figure 1: Category 1 Invalid Breakdown By Minute 3, 4, and 5
(Three Bases: Brooks, Kelly, and Randolph AFB [n=279])

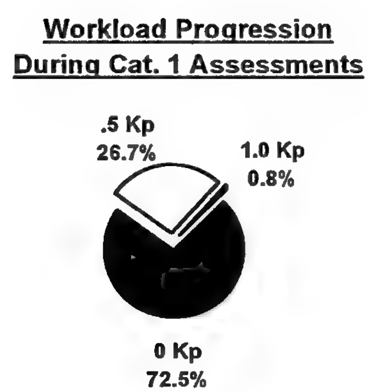
Minute Three:



Minute Four:



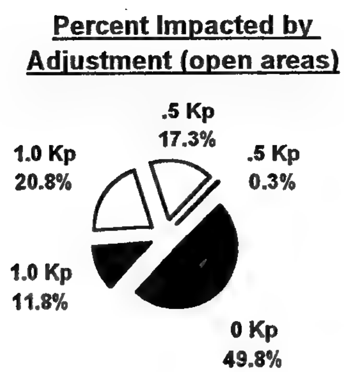
Minute Five:



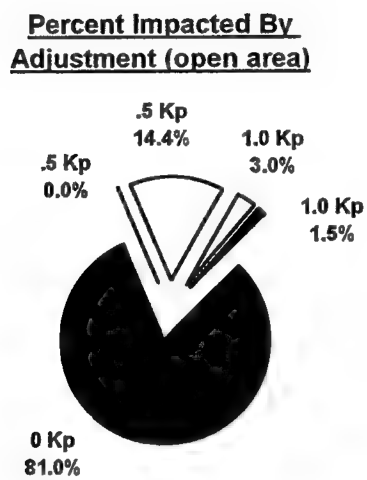
Legend: Open areas: Those that receive a WL increase.

Figure 2: Projected Impact On Category 1 Invalid Tests After A 10 Beat Per Minute Protocol Adjustment (n=279)

Minute Three:



Minute Four:



Minute Five:
No adjustment will be made to minute five.

Best Projected Impact:

- It is estimated that, at best, 55.5% of the Category 1 Invalid assessments can be affected (38.1% + 17.4% of Invalid assessments in minutes 3 and 4 from above).
- Category 1 Invalid assessments could be reduced to 22.7%of total Invalid assessments. This would reduce the number of Invalid assessments to 14.6% of all assessments taken.

Legend: Open areas: Those area impacted by 10 bpm Protocol adjustment.

TABLES

Table 1: Brooks, Kelly, Lackland, Patrick, and Randolph AFB Combined Cycle Ergometry Assessment Breakdown

Brooks, Kelly, Lackland, Patrick, and Randolph AFB Combined Assessment Results			
Test Result	# of assessments	% of Total Tests	
Invalids	1548	16.4	
Fail	904	9.6	
Pass	6985	74.0	
Total:	9437		
Test Result	# of assessments	% of Total Invalids	% of Total Tests
Invalid #1	615	39.7	6.5
Invalid #2	121	7.8	1.3
Invalid #3	233	15.1	2.5
Invalid #4	8	0.5	0.08
Invalid #5	137	8.9	1.5
Invalid #6	24	1.6	0.3
Invalid #7	410	26.5	4.3
Totals:	1548		

Table 1A: Brooks, Kelly, and Randolph AFB Combined Initial Cycle Ergometry Assessment Breakdown

Brooks, Kelly, and Randolph AFB Combined Assessment Results			
Test Result	# of assessments	% of Total Tests	
Invalids	660	16.2	
Fail	324	8.0	
Pass	3086	75.8	
Total:	4070		
Test Result	# of assessments	% of Total Invalids	% of Total Tests
Invalid #1	260	39.4	6.4
Invalid #2	21	3.2	0.5
Invalid #3	79	12.0	1.9
Invalid #4	2	0.3	0.05
Invalid #5	93	14.1	2.3
Invalid #6	7	1.1	0.2
Invalid #7	198	30.0	4.9
Totals:	660		

Table 2: Brooks, Kelly, and Randolph AFB Minute Three Workload Progression of Category 1 Invalid Assessments

	Brooks, Kelly, and Randolph AFB					
	Females		Males		Males and Females	
workload progression	# of assessments	% of total	# of assessments	% of total	# of assessments	% of total
1 Kp	14	20.9	77	36.3	91	32.6
.5 Kp	14	20.9	35	16.5	49	17.6
0 Kp	39	58.2	100	47.2	139	49.8
Total	67		212		279	

Table 3: Brooks, Kelly, and Randolph AFB Minute Four Workload Progression of Category 1 Invalid Assessments

	Brooks, Kelly, and Randolph AFB					
	Females		Males		Males and Females	
workload progression	# of assessments	% of total	# of assessments	% of total	# of assessments	% of total
1 Kp	1	1.6	11	5.5	12	4.5
.5 Kp	4	6.3	34	17	38	14.4
0 Kp	59	92.2	155	77.5	214	81.1
Total	64		200		264	

Table 4 : Brooks, Kelly, and Randolph AFB Minute Five Workload Progression of Category 1 Invalid Assessments

	Brooks, Kelly, and Randolph AFB					
	Females		Males		Males & Females	
workload progression	# of assessments	% of total	# of assessments	% of total	# of assessments	% of total
1 Kp	0	0	2	1	2	0.8
.5 Kp	8	13.1	60	30.9	68	26.7
0 Kp	53	86.9	132	68	185	72.5
Total	61		194		255	

Table 5: Brooks, Kelly, and Randolph AFB Category 1 Invalid Heart Rate Response During CE Assessment

Males and Females Combined							
Work Load Progression (WLP)	beats below initial workload inc.	Minute 3		Minute 4		Minute 5	
		# of assessments	% of total	# of assessments	% of total	# of assessments	% of total
1 Kp	1-5	34	12.2	6	2.3	1	0.3
	6-10	24	8.6	2	0.7	1	0.3
	>10	33	11.8	4	1.5	0	
.5 Kp	1-5	18	6.5	9	3.4	1	0.3
	6-10	30	10.8	29	11.0	6	2.4
	>10	1		0		61	23.9
	beats above lower limit of WLP						
0 Kp	1-5	7	2.5	25	9.5	14	5.5
	6-10	11	3.9	28	10.6	21	8.2
	11-15	9	3.2	27	10.2	26	10.2
	16-20	18	6.5	13	4.9	24	9.4
	>20	94	33.7	121	45.8	100	39.2
Total		279		264		255	

* Note: Lower limit of workload progression (i.e. highest HR at which an individual can receive a .5 Kp WL progression) determined by age and minute of progression (Appendix 1, Part A).

Table 6: Heart Rate Response During CE Assessment of 46 Individuals with Three Category 1 Invalid Assessments

Males and Females Combined							
Work Load Progression (WLP)	beats below initial workload inc.	Minute 3		Minute 4		Minute 5	
		# of assessments	% of total	# of assessments	% of total	# of assessments	% of total
1 Kp	1-5	4	6.8	0		0	
	6-10	2	3.4	1	1.7	0	
	>10	2	3.4	0		0	
.5 Kp	1-5	3	5.1	0		0	
	6-10	3	5.1	2	3.4	1	2
	>10	0		0		6	12
	beats above lower limit of WLP						
0 Kp	1-5	1	1.7	4	6.9	2	4
	6-10	3	5.1	4	6.9	1	2
	11-15	1	1.7	3	5.2	6	12
	16-20	8	13.6	4	6.9	6	12
	>20	32	54.2	40	69	28	56
Total		59		58		50	

* Note: Lower limit of workload progression (i.e. highest HR at which an individual can receive a .5 Kp WL progression) determined by age and minute of progression (Appendix 1, Part A).

Table 7: First and Second Re-Test Data for Individuals With an Initial Category 1, 2, 3, 4, or 6 Invalid Assessment

Brooks, Kelly, Lackland, Patrick, and Randolph AFB

First Re-Test	result	# of subjects	% of total assessments
	Pass	280	55.8
	Fail	111	22.1
	Invalid	111	22.1
	Total	502	
Second Re-Test	Pass	31	44.3
	Fail	14	20.0
	Invalid	25	35.7
	Total	70	

APPENDIX

Appendix 1

A) Heart rate parameters for workload progression (Protocol B and Original)

	Workload Progression											
	+1 kp			+0.5 kp			0.0 kp			Terminate Assessment		
Minute	3	4	5	3	4	5	3	4	5	3	4	5
Age												
17 - 30	<110	<110	<115	110-119	110-119	115-128	120-173	120-173	129-173	Invalid if >85% of max. heart rate		
31 - 40	<105	<105	<110	105-114	105-114	110-126	115-161	115-161	127-161			
41 - 50	<100	<100	<105	100-109	100-109	105-122	110-152	110-152	123-152			
51 - 60	<100	<100	<105	100-109	100-109	105-120	110-144	110-144	121-144			
61 - 70	<90	<90	<95	90-104	90-104	95-105	105-135	105-135	106-135			

Progression workload cycle changes*.

* Note: Heart rates used to determine workload progression are taken at the end of the minute. For example, minute three of the assessment is performed at the initial workload, with the heart rate at the end of minute three determining the workload progression for minute four using "Minute 3" workload progression column.

B) Heart rate parameters for workload progression (Protocol A)

	Workload Progression											
	+1 kp			+0.5 kp			0.0 kp			Terminate Assessment		
Minute	3	4	5	3	4	5	3	4	5	3	4	5
Age												
17 - 30	<100	<100	<115	100-119	100-119	115-128	120-173	120-173	129-173	Invalid if >85% of max. heart rate		
31 - 40	<95	<95	<110	95-114	95-114	110-126	115-161	115-161	127-161			
41 - 50	<90	<90	<105	90-109	90-109	105-122	110-152	110-152	123-152			
51 - 60	<90	<90	<105	90-109	90-109	105-120	110-144	110-144	121-144			
61 - 70	<80	<80	<95	80-104	80-104	95-105	105-135	105-135	106-135			

Progression workload cycle changes.

A TEST OF THREE MODELS OF THE ROLE OF g AND PRIOR JOB KNOWLEDGE IN THE ACQUISITION OF SUBSEQUENT JOB KNOWLEDGE

Thomas W. Doub
Graduate Student
Department of Psychology and Human Development

Vanderbilt University
18th and Horton Ave.
Nashville, TN 37203

Final Report for:
Summer Graduate Student Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Armstrong Laboratory Human Resources Directorate

August 1996

A TEST OF THREE MODELS OF THE ROLE OF g AND PRIOR JOB KNOWLEDGE IN THE ACQUISITION OF SUBSEQUENT JOB KNOWLEDGE

Thomas W. Doub
Graduate Student
Department of Psychology and Human Development
Vanderbilt University

Abstract

Based on data from 83 independent studies with a total sample of 42,399 participants, structural equation models were used to test three theories of the role of ability and prior job knowledge on the acquisition of subsequent job knowledge. Ability and prior job knowledge were measured before entering job training, and subsequent job knowledge was measured at the completion of job training. The three models were: a) a role for ability only, b) a role for prior job knowledge only, and c) a role for both ability and prior job knowledge. Results supported the model with a role for both ability and prior job knowledge. The R^2 for predicting subsequent job knowledge for the model including all jobs was .80. Three other analyses were conducted within job families with very similar results. In all analyses, the causal impact of ability was far greater than the causal impact of prior job knowledge. In the model considering all jobs, the causal impact of ability was about three times that of prior job knowledge.

A Test of Three Models of the Role of g and Prior Job Knowledge
in the Acquisition of Subsequent Job Knowledge

The role of general cognitive ability (g) in influencing job performance has been well demonstrated. Hunter (1983, 1986) has shown in meta-analytic path analyses that the major way ability influences job performance is through the acquisition of job knowledge. He demonstrated a strong causal path between ability and job knowledge in samples of both military and civilian non-supervisory job incumbents for jobs of low to moderate complexity. Ree, Carretta, and Teachout (1995) extended these findings in a military training-based study of the acquisition of aircraft pilot knowledge and skills, a high complexity job. Borman, Hanson, Oppler, Pulakos, and White (1993) further extended Hunter's model to demonstrate that ability influenced supervisory job knowledge as well. Schmidt, Hunter, and Outerbridge (1986) and Borman and associates (Borman, White, & Dorsey, 1995; Borman, White, Pulakos, & Oppler, 1991) have further corroborated the relationship between ability and job knowledge. The following causal relationship has been widely accepted:

ability → job knowledge → job performance.

Ability leads to job knowledge and job knowledge applied, leads to job performance.

Dye, Reck, and McDaniel (1993) have shown the validity of job knowledge tests as predictors of job performance in a meta-analysis. They found average validity of about .47 for job training criteria and a similar validity of .45 for job performance criteria. Ree et al. (1995) postulated prior job knowledge as job-relevant knowledge the employee possesses prior to beginning the job or job training. They tested the role of g and prior job knowledge in training and found that g had a strong causal influence on the acquisition of prior job knowledge and subsequent job knowledge with standardized path coefficients of .62. However, prior job knowledge, in turn, had a weak, .02, causal influence on the acquisition of additional job knowledge acquired during training.

Models of the Role of g and Prior Job Knowledge

Three models of g, prior job knowledge, and the acquisition of subsequent job knowledge are suggested by various psychological theories. The first model, g-only, specifies that g is the only influence on the acquisition of subsequent job knowledge and that prior job knowledge exerts no causal impact. This is the model often attributed to researchers who find value in g as a predictor or causal variable. For example, Sternberg and Wagner (1993) stated with regards to what they called the "g-ocentric model"

> One of the things that is predicted by g, according to g-ocentric theorists, is job performance. In this view, if an employer were to use only intelligence tests to select the highest scoring applicant for each job, training results would be predicted well regardless of the job, and overall performance from the employees selected would be maximized. (p. 1)

The second model, prior-job-knowledge-only, specifies that g has no influence and that only prior job knowledge influences the acquisition of subsequent job knowledge. This model was based on authors' interpretation

of studies of chess players (Chase & Simon, 1973; de Groot, 1965), horse race handicapping and betting (Ceci & Liker, 1986), practical intelligence (Sternberg & Wagner, 1993), and competency testing theory (McClelland, 1973). See Perkins (1995) for a broad review of what he terms "experiential intelligence" including such tasks as computer programming, solving physics problems, and others.

The third is a joint model that specifies positive causal path for both g and prior job knowledge in the acquisition of subsequent job knowledge. This model was based on the findings of Hunter (1983, 1986) and Dye et al. (1993). A broadly based test of the joint model has not been conducted.

These three models allow predictions to be made about the relative magnitude of standardized path coefficients. Figures 1a through 1c show the expected valence of the coefficients associated with each of the models.

Insert Figure 1 About Here

The g-only model, Figure 1a, would yield a zero causal path between prior job knowledge and subsequent job knowledge, but a positive causal path between g and subsequent job knowledge. Conversely, the prior-job-knowledge-only model hypothesizes a positive causal path from prior job knowledge to subsequent job knowledge and zero paths from g (Figure 1b). Finally, Figure 1c shows the joint model hypothesizing positive causal paths for both g and prior job knowledge to subsequent job knowledge.

The purpose of this study was to evaluate these three models across multiple jobs in two broad job families. This extends our cumulative understanding of the role of g and prior job knowledge on the acquisition of subsequent job knowledge acquired during training.

Method

Participants

The participants were 42,399 United States Air Force enlisted men and women who had attended and completed a technical training course for one of several job specialties in either the electronics or mechanical career field. They were between the ages of about 17 and 23, predominantly male (83%) and White (80%), with a high school or better education (99%). All had tested for enlistment qualification between 1984 and 1988. Participants for these jobs were selected on the basis of both g and prior job knowledge before entering extensive formal technical training.

The sample sizes were: electronics, 32,140; mechanical, 19,289; electronics/mechanical, 9,030; all training, 42,399. There is overlap between the electronics and mechanical samples because of entry requirements for training.

Some training courses require applicants to qualify on both electronics and mechanical composites, and some allow qualification on either the electronics or mechanical knowledge composites.

Measures

g and prior job knowledge. General cognitive ability, g, and prior job knowledge, (JK_P), were measured simultaneously with the Armed Services Vocational Aptitude Battery (ASVAB; Earles & Ree, 1992). The ASVAB is developed from a detailed written taxonomy that specifies both content and psychometric characteristics.

The ASVAB consists of 10 subtests that measure g and lower-order factors of verbal/quantitative, technical job knowledge, and speed (Ree & Carretta, 1994). The verbal and quantitative subtests are Word Knowledge (WK), Paragraph Comprehension (PC), Arithmetic Reasoning (AR), and Mathematics Knowledge (MK). Electronics Information (EI), Mechanical Comprehension (MC), Auto and Shop Information (A/S), and General Science (GS) are the technical knowledge subtests. The two speed subtests are Numerical Operations (NO) and Coding Speed (CS).

For the purposes of this study, g was extracted as a latent factor from the two verbal and two quantitative subtests (WK, PC, AR, and MK), generally accepted measures of g. The verbal subtests are measures of synonyms (WK) and short-paragraph reading comprehension (PC), while the quantitative subtests are measures involving word problems (AR) and problem solving using high school mathematics (MK).

Job knowledge was extracted from EI for electronics jobs and MC and A/S for mechanical jobs. The GS, NO, and CS subtests were not used because they do not measure job knowledge that is specific to any job family.

The EI subtest is a measure of knowledge about elementary electrical principles and electronics (e.g., "Which schematic symbol indicates a resistor?"). The A/S subtest measures knowledge about shop practices (e.g., "What is the tool pictured above?") and about automotive systems (e.g., "In which system is an EGR valve found?"). MC measures knowledge of mechanical principals and tools (e.g., "In this arrangement of pulleys, which pulley turns fastest?").

The following describes subtests that are not electronics or mechanical job knowledge measures. The GS subtest is a measure of knowledge of biology, earth science, and elementary physical science (e.g., "The lack of iodine is often related to which of the following diseases?" "The mantel describes which layer of the earth?" "Water is an example of which state of matter?"). The NO subtest is a series of 50 arithmetically trivial items (e.g., 2 +3 = 5, 60 / 4 = 15, 3 x 4 = 12, 6 - 5 = 1) that must be completed in three minutes. The CS subtest requires the examinee to find the number that goes with specific words from a table. There are 84 CS items that must be answered in 11 minutes. NO and CS are the least g-saturated tests in the battery. The ASVAB Information Pamphlet (DOD, 1984) given to all applicants prior to testing provides example items for each subtest.

Criteria. The criteria, subsequent job knowledge acquired in training, were final grades on tests of job knowledge earned by the participants during technical training. These grades ranged from 70 to 99 and were the average percent correct on several (at least four, but sometimes more) multiple choice tests. Each course scaled the

grades independently and no common metric exists for the set of grades. The reliability of the training grades was not estimated by those assigning grades nor were the data available to estimate reliability directly.

The training courses typically last between two and eight months depending on the job specialty. Attrition rates for these courses are quite low with an average of six percent. Attrition has several characteristics in military training. Some who fail are separated from service and some are transferred to other training or to jobs not requiring formal training.

Job Families

The Air Force aggregates all jobs into one of four major job families. These job families are the result of clustering regression equations of the ASVAB subtests (Alley, Treat, & Black, 1988) and policy decisions by senior executives. The regressions all used final technical training grades as criteria. The Air Force uses both general ability (i.e., AFQT or Armed Forces Qualification Test) and specific composites (Mechanical, Administrative, General, and Electronics) of the multiple aptitude battery for placing applicants into specific jobs.

All applicants are screened on g via the composite of two verbal and two quantitative tests (i.e., AFQT). In addition to the scores on the g composite for jobs in the electronics family, minimum scores on a composite composed of Electronics = AR + MK + EI + GS must be achieved. EI is a measure of prior job knowledge for electronics jobs. Applicants for mechanical jobs must qualify on a composite made up of Mechanical = MC + GS + 2A/S. MC and A/S are measures of prior job knowledge for mechanical jobs. Even though GS is a measure of technical knowledge, it is not a content-relevant measure of job knowledge for these occupations.

A small number of jobs allow the applicant to qualify on either of these two composites. The only composites containing job knowledge tests used in this way are Electronics and Mechanical. Some jobs allow qualification "if Electronics is greater than the 30th percentile or Mechanical is greater than the 35th percentile." Qualification on both is sometimes required (i.e., a minimum percentile of 30 on Electronics and a minimum percentile of 47 on Mechanical).

Electronics jobs include the broad areas of precision measurement equipment repair and calibration, communications-electronics repair and maintenance, aircraft electronics, missile electronics maintenance, and others. Similarly, mechanical jobs include the broad areas of missile, vehicle, and airframe maintenance, munitions and weapons, structural/pavements, fuels, and others. The assignment of jobs to job families and minimum test score requirements are controlled by official regulations.

Neither of the other two job families (i.e., administrative and general) use measures of job knowledge. Jobs in the administrative family in addition to qualification on the g composite, require minimum scores on a composite of Administrative = WK + PC + NO + CS. Although, NO and CS create a specific speed factor (Ree & Carretta, 1994) they are not measures of job knowledge. Representative jobs in the administrative job family are supply, personnel clerk, financial management, and transportation administration. The jobs in the general family, in addition to qualification on the g composite, require minimum scores on a composite of three (General = AR + WK + PC) of

the four subtests in the g-composite. Representative jobs in the general job family are intelligence, fire protection, medical technology, and law enforcement.

Data Analyses

Because the participants were selected, at least in part, on the basis of the subtest scores, they represented a range restricted sample. To correct for the estimation bias introduced by range restriction, we individually corrected each job-specific correlation matrix of subtest scores and criterion. Eighty-three individual matrices were corrected to reflect unrestricted correlations in the normative sample of the aptitude battery (Bock & Moore, 1984; Ree & Wegner, 1990). The multivariate procedure of Lawley (1943) was used.

After correcting for range restriction, the correlations for all 10 subtests and the criterion for each job were averaged within each job family and across job families. That is, there was an average correlation between each pair of subtests and between each subtest and the criterion. These were sample-weighted averages.

We followed the method suggested by Viswesvaran and Ones (1995) for combining meta-analyses and path models. The sample-weighted average correlations provide better estimates of the true relationships than do individual correlations. In this study, path analyses of these meta correlations are superior to ordinary regression analyses because path analyses require an explicit causal approach to the explanation of phenomena (Asher, 1976).

Structural equation analyses based on the weighted-averaged correlations were estimated with the LISREL 8 program (Jöreskog & Sörbom, 1993). There was a mixture of latent and observed variables in all measurement models.

The measure of g was a latent variable derived from the two verbal and two quantitative subtests. An analysis was conducted to evaluate the magnitude of the Eigenvalues. A relatively large first Eigenvalue would be consistent with a general factor. Additionally, a confirmatory factor analysis of the four subtests was done to assess the fit of a single factor model.

For electronics jobs, prior job knowledge was the electronics information subtest score, an observed variable. For mechanical jobs, prior job knowledge was a latent variable derived from the mechanical comprehension and auto and shop information tests. For the model containing both mechanical and electronic jobs and the model where qualification was based on either electronics or mechanical or both, prior job knowledge was a latent variable derived from mechanical comprehension, auto and shop information, and electronics information subtest scores. The criterion, subsequent job knowledge (JK_S), was an observed variable in all models. For the observed variable electronics information, the reliability (.80) was taken from Bock and Moore (1984). In all models, the reliability (.80) of the observed criterion variables was taken from Pearlman, Schmidt, and Hunter (1980). These values were used in the structural equation analysis.

First, separate models for electronics jobs and mechanical jobs were constructed. Then a model was constructed for jobs that allowed qualification on both electronics and mechanical or either electronics or mechanical. Finally, a model using all 83 jobs was constructed.

Both the direct and indirect influence for each antecedent variable was calculated as was the R^2 for the dependent variable of subsequent job knowledge. The goodness-of-fit for the structural model was measured by the Comparative Fit Index (CFI) and Root Mean Square Error of Approximation (RMSEA). The CFI is an extension of the Tucker-Lewis fit index but not as sensitive to sample size. Values above .90 are considered as good fit. RMSEA is a measure of error per parameter estimated. The lower the RMSEA the better.

Results

Table 1 shows the correlations of the subtests and criteria. The first seven variables are the subtests and the last four variables are the criterion measures. Note that the corrected intercorrelations of the subtests are the same for all models. This is because the correction for range restriction for each job family used the same normative population. However, the correlations between the subtest scores and the criterion vary by job family. There are no correlations among the criterion variables, as none of the participants completed more than one job training course.

Insert Table 1 About Here

An Eigenanalysis of the four subtests used to measure g disclosed one large value of 3.1, accounting for 79 percent of the variance. Minor verbal and quantitative content factors accounted for 9, 7, and 4 percent of the variance. This common variance has been defined as g elsewhere (Jensen, 1980) and scores from these subtests have been used as measures of g elsewhere (Herrnstein & Murray, 1994). Figure 2 shows the loadings of the subtests on the common factor as estimated by the confirmatory factor analysis. The loadings of the subtests on the common factor were .76, .81, .83, and .89. The CFI was 1.00, with an RMSEA of .04, indicating a good fit for the single factor model.

Insert Figure 2 About Here

Table 2 shows the correlations among the latent variables as estimated in the measurement model by LISREL 8 for electronics jobs, mechanical jobs, electronics and/or mechanical jobs, and all jobs.

Insert Table 2 About Here

Insert Figure 3 About Here

Figure 3 shows the path models and their standardized path coefficients. The path coefficient between g and prior job knowledge was .86 for the model with the electronics jobs. The total direct and indirect contribution of g to subsequent job knowledge was .89. The contribution of prior job knowledge to subsequent job knowledge was .26. The causal impact of g was 3.42 times that of prior job knowledge. The R^2 for predicting the criterion of subsequent job knowledge was .82. Measures of model fit were in an acceptable range (CFI = .99, RMSEA = .09).

In order to improve the fit of all these models, especially RMSEA, paths between the observed variables comprising g and the observed variables comprising prior and subsequent job knowledge could be freed. However, since the composition of prior and subsequent job knowledge varies at the measurement level across job families, these paths were fixed at zero to maintain comparability across models.

For the model with the mechanical jobs, the path coefficient between g and prior job knowledge was .80. The total contribution of g to subsequent job knowledge was .80, counting both direct and indirect paths. The contribution of prior job knowledge to subsequent job knowledge was .26. The causal impact of g was 3.08 times that of prior job knowledge. The R^2 for predicting the criterion of subsequent job knowledge was .73. Fit for the mechanical model was acceptable (CFI = .97, RMSEA = .12).

Much the same was found for occupations allowing qualification on both or either electronics and mechanical. The path coefficient between g and prior job knowledge was .84 and the total direct and indirect contribution of g to subsequent job knowledge was .86. The direct effect of prior job knowledge on subsequent job knowledge was .36. The causal impact of g was 2.38 times that of prior job knowledge. The R^2 for predicting the criterion of subsequent job knowledge was .78. Fit for the electronics/mechanical model was acceptable (CFI = .96, RMSEA = .13).

The path coefficient between g and prior job knowledge for the model with all jobs combined was .84. The total causal effect of g on subsequent job knowledge was .88, counting both direct and indirect paths. The causal effect of prior job knowledge on subsequent job knowledge was .29. The causal effect of g was 3.03 times that of prior job knowledge. Further, the R^2 for predicting the criterion of subsequent job knowledge was .80. Fit for the model including all jobs was again acceptable (CFI = .96, RMSEA = .13).

Discussion

Neither the g-only model nor the prior-job-knowledge-only model was supported. The g-only model was rejected because there was a non-zero path between prior job knowledge and subsequent job knowledge. Also, there was a non-zero path coefficient between g and subsequent job knowledge, which rejected the prior-job-knowledge-only model. The mixed model was supported in all analyses.

There are no advocates for the g-only model. This is the straw man erected by critics of general cognitive ability (for example, Sternberg & Wagner, 1993), not a model offered by g researchers. The lack of support for the prior-job-knowledge-only model is consistent with Barrett and Depinet's (1991) reconsideration of McClelland's (1973) claims about testing for competency rather than for intelligence. It is inconsistent with de Groot (1965), Chase and Simon, (1973), and Ceci and Liker's (1986) claims about special knowledge. Perhaps the work of these authors should be viewed with structural models that take into account the role of g in the acquisition of job knowledge.

Job family did not moderate the causal relationships among the latent variables. The differences in the path coefficients were quite small across job families, never exceeding .10. The smallest difference, .06, was found for the path from g to prior job knowledge. The path coefficients are remarkably consistent across models, indicating that the effects of g and prior job knowledge are generally equivalent across job families.

This similarity across models is quite striking given substantial differences in the respective measurement models. Since the measures used in each structural model were specific to a given job family, differences at the level of the latent variables might have been expected. However, differences were not observed. Also, the number of subtests used in the estimation of the prior job knowledge latent variable differed by job family. Additional subtests could presumably provide more comprehensive coverage of the construct and more reliable measurement. However, in these analyses, the number of subtests did not appear to affect the relationship among latent variables.

The direct effect of g on subsequent job knowledge was twice that of prior job knowledge. The total (direct and indirect) effect of g on subsequent job knowledge was about three times that of prior job knowledge indicating that g plays a more important role in the acquisition of subsequent job knowledge. Although these analyses are set in the context of industrial psychology, it is expected that the relative importance of g in the learning process would generalize to a variety of settings, most notably education.

The model, as indicated by the R^2 values, explains approximately 80% of the variation in subsequent job knowledge. This leaves approximately 20% of the variance unaccounted for. The work of Borman and associates (1991, 1995) suggests that the personality construct of conscientiousness would account for a proportion of the unexplained variance. Based on Borman et al. we would expect a standardized path coefficient between conscientiousness and subsequent job knowledge of about .13. Again, based on Borman et al. (1995), an increase in R^2 of about .02 would be expected for adding conscientiousness to the model with g and prior job knowledge. Finally, we speculate that the role of interest or motivation may be manifested in the scores on these prior job knowledge tests. Subject participants may have acquired knowledge on the basis of personal choice to attend classes or through reading outside of formal education. This would also contribute to the substantial R^2s observed here.

The R^2 values in the current study were greater than those reported in similar studies (Borman et al., 1991, 1993, 1995; Hunter, 1983, 1986; Schmidt et al., 1986). There may be two reasons for this. First, the construct of

prior job knowledge was added. This additional variable allows, but does not guarantee, better prediction. Second, the correlations were of latent variables corrected for range restriction. Therefore, values around .80 are not surprising given knowledge of the causal impact of g on training performance.

Ree et al. (1995) reported a path coefficient between g and prior job knowledge of .62 for the occupational category of aircraft pilot. In the current study, the path coefficient between g and prior job knowledge for the model with all jobs was .84. The .84 value is the consequence of results from 83 jobs based on 42,399 participants, whereas the .62 coefficient from Ree et al. is based on one job and 3,428 participants.

Although the jobs spanned a broad range of electronic and mechanical occupations, the mixed model was supported by all results. Job family did not moderate the relationships among the latent variables. Though both g and prior job knowledge were found to have a causal impact on subsequent job knowledge, the relative impact of g was three times that of prior job knowledge.

References

Alley, W. E., Treat, B. R., Black, D. E. (1988). Classification of Air Force jobs into clusters (AFHRL-TR-88-14). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Asher, H. B. (1976). Causal Modeling. In E. M. Uslaner (Series Ed.), Quantitative applications in the social sciences. Beverly Hills, CA: Sage.

Barrett, G. B., & Depinet, R. L. (1991). A reconsideration of testing for competence rather than for intelligence. American Psychologist, 46, 1012-1024.

Bock, R. D., & Moore, E. G. J. (1984). The profile of American youth: Demographic influences on test performance. Washington DC: GPO.

Borman, W. C., Hanson, M. A., Oppler, S. H., Pulakos, E. D., & White, L. A. (1993). Role of early supervisory experience in supervisor performance. Journal of Applied Psychology, 78, 443-449.

Borman, W. C., White, L. A., & Dorsey, D. W. (1995). Effects of ratee task performance and interpersonal factors on supervisor and peer performance ratings. Journal of Applied Psychology, 80, 168-177.

Borman, W. C., White, L. A., Pulakos, E. D., & Oppler, S. H. (1991). Models of supervisory job performance ratings. Journal of Applied Psychology, 76, 863-872.

Ceci, S. J., & Liker, J. (1986). A day at the races: A study of IQ, expertise, and cognitive complexity. Journal of Developmental Psychology: General, 115, 255-266.

Chase, W. C., & Simon, H. A. (1973). Perception in chess. Cognitive Psychology, 4, 55-81.

de Groot, A. D. (1965). Thought and choice in chess. The Hague: Mouton.

Department of Defense (1984). Armed Services Vocational Aptitude Battery (ASVAB) Information Pamphlet (DOD 1304.12Z).North Chicago, IL: United States Military Enlistment Processing Command.

Dye, D. A., Reck, M., & McDaniel, M. (1993). The validity of job knowledge measures. International Journal of Selection and Assessment, 1, 153-162.

Earles, J. A., & Ree, M. J. (1992). The predictive validity of ASVAB for training grades. Educational and Psychological Measurement, 52, 721-725.

Herrnstein, R., & Murray, C. (1994). The bell curve. NY: The Free Press

Hunter, J. E. (1983). A causal analysis of cognitive ability, job knowledge, job performance, and supervisory ratings. In F. Landy, S. Zedeck, & J. Cleveland (Eds.). Performance measurement and theory (pp. 257-266). Hillsdale, NJ: Erlbaum.

Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. Journal of Vocational Behavior, 29, 340-362.

Jensen, A. R. (1980). Bias in mental testing. NY: The Free Press.

Jöreskog, K. G. & Sörbom, D. (1993). LISREL 8 user's reference guide. Chicago, IL: Scientific Software International.

Lawley, D. N. (1943). A note on Karl Pearson's selection formulae. Proceedings of the Royal Society of Edinburgh, Section A, 62, Part I, 20-30.

McClelland, D. C. (1973). Testing for competence rather than for "intelligence." American Psychologist, 28, 1-14.

Pearlman, K, Schmidt, F. L., & Hunter, J. E. (1980). Validity generalization results for tests used to predict job proficiency and training success in clerical occupations. Journal of Applied Psychology, 65, 373-406.

Perkins, D. (1995). Outsmarting IQ: The emerging science of learnable intelligence. NY: The Free Press.

Ree, M. J., & Carretta, T. R. (1994). Factor analysis of the ASVAB: Confirming a Vernon-like structure. Educational and Psychological Measurement, 54, 459-463.

Ree, M. J., Carretta, T. R., & Teachout, M. S. (1995). Role of ability and prior job knowledge in complex training performance. Journal of Applied Psychology, 80, 721-730.

Ree, M. J., & Wegner, T. G. (1990) Correcting differences in Answer sheets for the 1980 Armed Services Vocational Aptitude Battery reference population. Military Psychology, 2, 157-169.

Schmidt, F. L., Hunter, J. E., & Outerbridge, A. N. (1986). Impact of job experience and ability on job knowledge, work sample performance, and supervisory ratings of job performance. Journal of Applied Psychology, 71, 432-439.

Sternberg, R. J., & Wagner, R. K. (1993). The g-ocentric view of intelligence and job performance is wrong. Current Directions in Psychological Science, 2, 1-5.

Viswesvaran, C., & Ones, D. S. (1995). Theory testing: Combining psychometric meta-analysis and structural equations modeling. Personnel Psychology, 48, 865-885.

Table 1.

Correlations of the Variables

	AR	WK	PC	A/S	MK	MC	EI	E	M	E/M	ALL
AR	1.000										
WK	.722	1.000									
PC	.672	.803	1.000								
A/S	.533	.529	.423	1.000							
MK	.827	.670	.637	.415	1.000						
MC	.693	.593	.593	.741	.600	1.000					
EI	.658	.684	.573	.745	.585	.743	1.000				
E	.699	.654	.607	.594	.664	.649	.669	1.000			
M	.660	.620	.580	.590	.610	.630	.640	---	1.000		
E/M	.671	.638	.601	.610	.626	.641	.659	---	---	1.000	
ALL	.688	.644	.597	.590	.649	.640	.658	---	---	---	1.000

Note. The "---" indicates no correlations between criterion variables could be computed because each participant was in only one training course. E, M, E/M, and ALL stand for the criteria for electronics jobs, mechanical jobs, jobs requiring electronics and/or mechanical, and all jobs.

Table 2.

Correlations among the latent variables

Electronics Jobs

g	JK_P	JK_S
1.000		
.862	1.000	
.894	.836	1.000

Mechanical Jobs

g	JK_P	JK_S
1.000		
.798	1.000	
.839	.763	1.000

Electronics and/or Mechanical Jobs

g	JK_P	JK_S
1.000		
.836	1.000	
.863	.828	1.000

All Jobs

g	JK_P	JK_S
1.000		
.835	1.000	
.878	.821	1.000

Note. Correlations were estimated by the structural equation program. JK_P and JK_S are prior job knowledge and subsequent job knowledge acquired during training, respectively.

Figure 1. Hypothesized path models for the three theories of job knowledge acquisition.

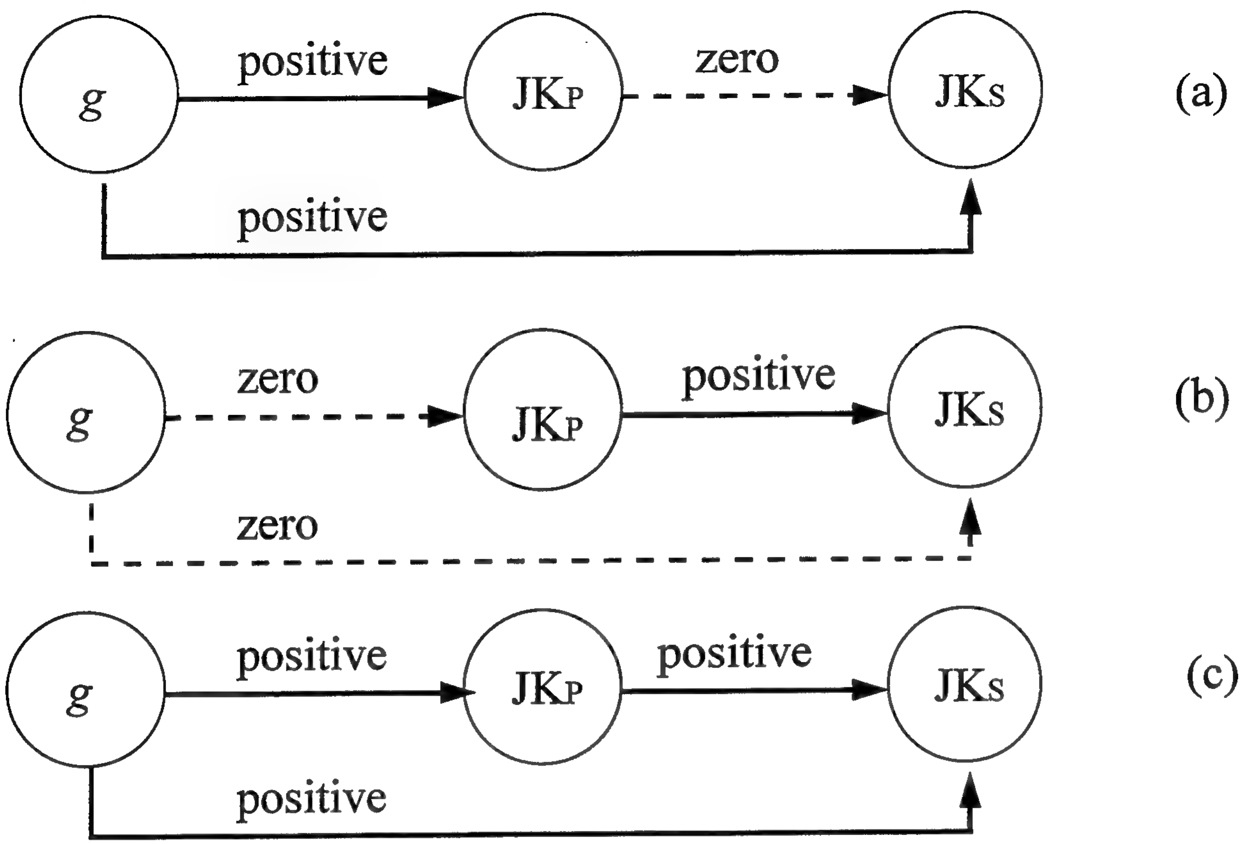


Figure 2. Confirmatory factor analysis of cognitive ability tests.

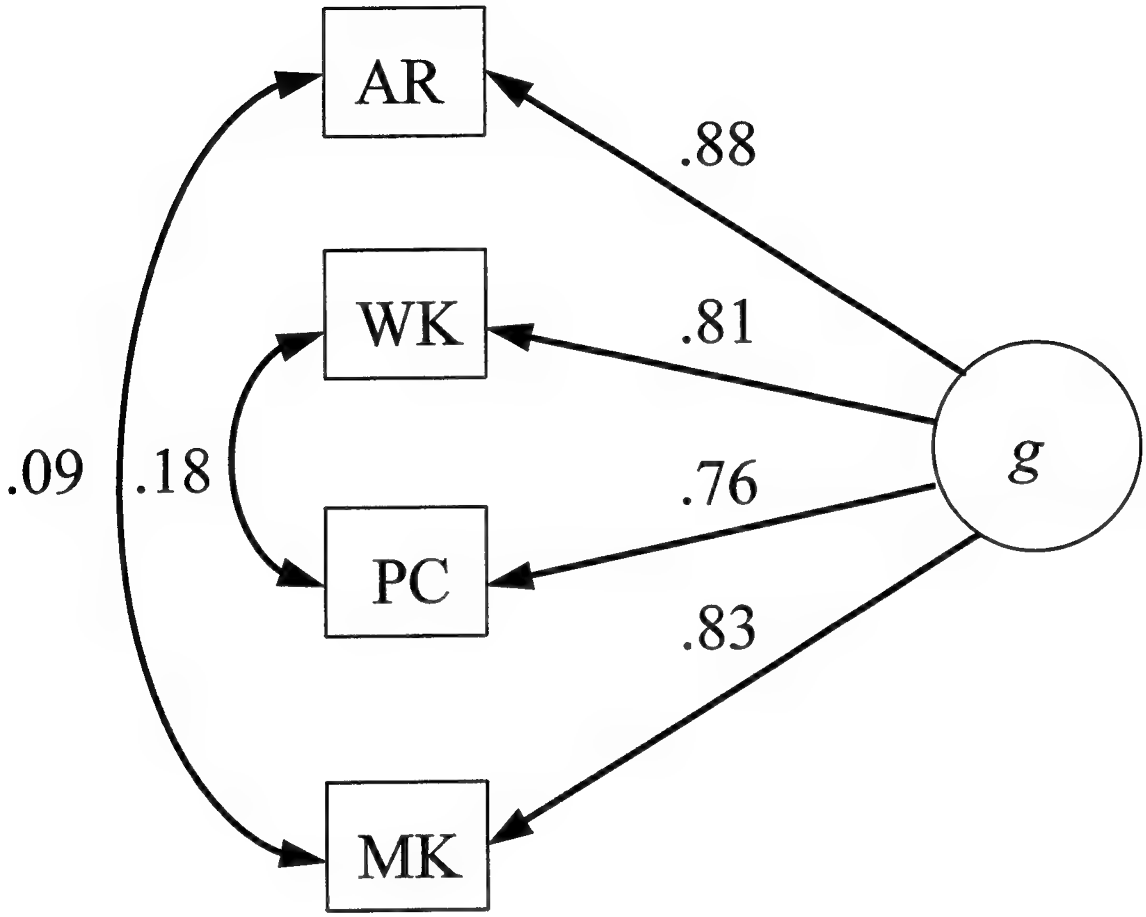
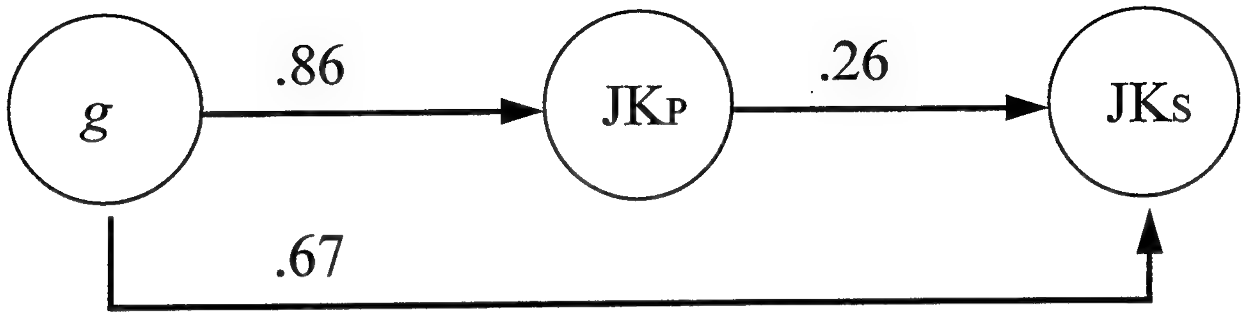
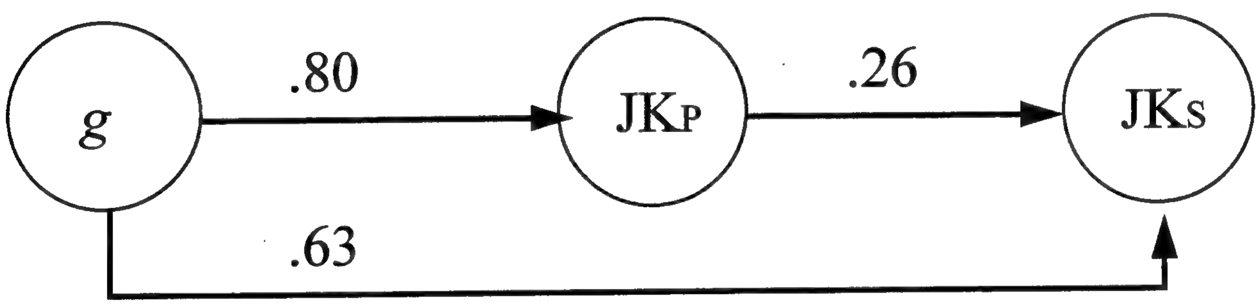


Figure 3. Path models for job families.

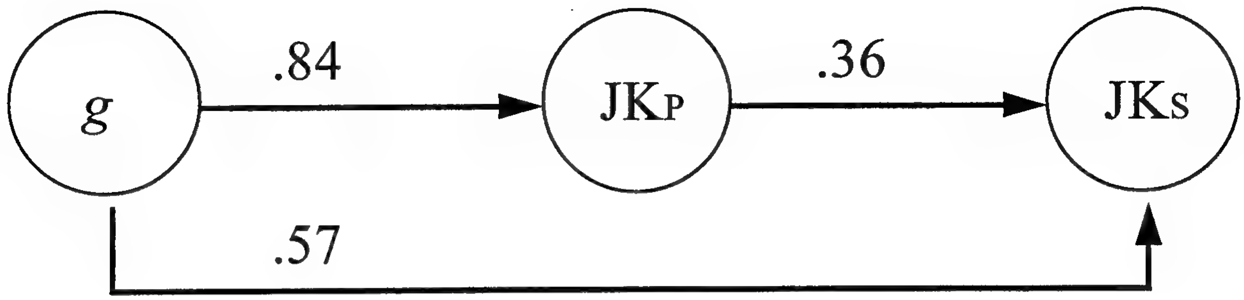
58 Electronics Jobs



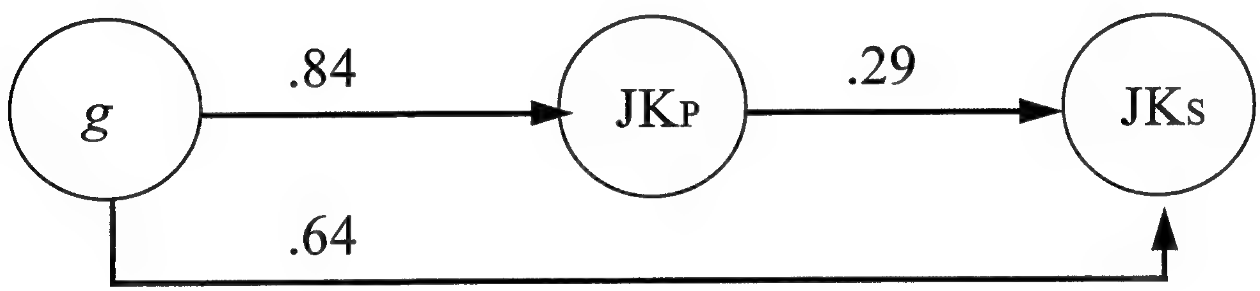
39 Mechanical Jobs



14 Electronics and/or Mechanical Jobs



All 83 Electronics and Mechanical Jobs



TIME-TO-CONTACT JUDGMENTS IN THE PRESENCE OF STATIC AND DYNAMIC OBJECTS: A PRELIMINARY REPORT

Philip H. Marshall, Professor
Ronald D. Dunlap, Doctoral Candidate
Department of Psychology

Texas Tech University
Lubbock, TX 79409-2051

Final Report for:
Summer Faculty Research Program
and
Graduate Student Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Armstrong Laboratory

August 1996

TIME-TO-CONTACT JUDGMENTS IN THE PRESENCE OF STATIC AND DYNAMIC OBJECTS: A PRELIMINARY REPORT

Philip H. Marshall, Professor
Ronald D. Dunlap, Doctoral Candidate
Department of Psychology
Texas Tech University

Abstract

The accuracy of time-to-contact (*TTC*) judgments in computer-generated visual displays was investigated in conditions that included no, static, or dynamic (moving) non-target stimuli. The number of such stimuli, and their direction and relative speed of movement also were manipulated. Analyses indicated that our tasks yielded traditional *TTC* functions, with undersestimation increasing as actual *TTC* increased (2-, 4-, 8-sec). The direction of non-target stimuli movement influenced *TTC* judgments only when they traveled at the same speed and in the same direction as the target. This effect was most pronounced at the longest *TTC*. Neither the number of non-target stimuli, nor non-target movement in general, affected *TTC* estimates. We suggest that a non-target stimulus may play several roles (have several influences) depending on the task requirements and the display configuration. Ordinarily one would think of non-target stimuli as distractors, but we suggest that when a non-target stimulus moves in the same direction and at the same speed as a target, it can assume the role of a "surrogate target," providing visible cues with which to judge target *TTC*. Within the limits of the conditions of this study, we conclude that *TTC* estimates are very robust, and are not easily influenced by otherwise extraneous variables, including accidental and potentially adverse testing environments. Performance on a *TTC* task, however, also may be determined by the adaptive nature of general strategic cognitive processes. We propose further research to determine if, when, and how extraneous stimuli may influence *TTC* accuracy, and what other adaptive and non-automatic processes might be inloved.

TIME-TO-CONTACT JUDGMENTS IN THE PRESENCE OF STATIC AND DYNAMIC OBJECTS: A PRELIMINARY REPORT

Philip H. Marshall and Ronald D. Dunlap

Introduction

For some time there has been a research interest in the ability of human observers to make time-to-contact (*TTC*) judgments. In one common version of this task, an observer watches a target traveling horizontally (at constant velocity) along a path for several seconds before that target disappears. The participant is to predict (usually by pressing a button) when the target would reach a predetermined end point or finish line. Typically, performance is characterized by increasing underestimation of *TTC* (responding earlier than the target would have made contact) as actual *TTC* increases (Schiff & Detwiler, 1979; Caird & Hancock, 1991). Some esearchers have suggested this ability to be solely a function of information from the optic array (Lee, 1976; Tresilian, 1991), while others have suggested the involvement of various cognitive processes and mechanism such as memory, imagery, and internal clocks (see Tresilian, 1995).

The stimuli in most *TTC* tasks consist of simple, moving objects (e.g., a square) in uncluttered displays, with no other stimuli. There are attempts currently underway to assess some potential distractor effects (Jennifer Blume, March 6, 1996; Gregory Liddell, May 6, 1996), and one published study (Lyon & Wagg, 1995) reports limited non-target stimulus effects with a target moving in a circular path. Research incorporating potentially distracting or other stimuli in the visual field can make contributions in several ways. First, real world situations in which *TTC* judgments are made are very likely to contain distracting or other events, and this is so even if the "real world" task is only monitoring a computer display. Therefore, research incorporating non-target stimuli is somewhat more "ecologically valid" than that where only a target is present and moves. Such research could also contribute to the debate on the extent of involvement of cognitive processes in *TTC* decision tasks. Cognitive acts that require effort (as distinguished from those that have become automatic) require a share of our limited attentional resources. To the extent that *TTC* processing is effortful, sufficiently distracting events could reduce attentional resources and affect *TTC* performance. Alternatively, there are other perceptual phenomena that might affect *TTC* accuracy when other stimuli, especially moving stimuli, are present in the visual array, and an example would be the so-called motion repulsion effect described by Marshak and Sekuler (1979). They found that the

perceived direction of motion for a given dot can be affected by the motion of another dot in the visual array such that the perceived difference between their respective headings is exaggerated.

In the present study, the presence, number, and direction of moving non-target stimuli were manipulated to determine possible effects on the perception of either the target's speed or path that would affect the accuracy of *TTC* judgments. It is worth noting that the nature of the effects of non-target stimuli could be to move the *TTC* function closer to actual times, that is, compensate for the underestimation normally observed. So, it would be naive to assume that the effects of the presence of non-target stimuli should always be in the direction of decreased performance, and we recognize that stimuli may have various functional roles depending on the situations in which they are present.

Method

Design

The variety of trials (stimulus scenes) in this study included those on which no non-target stimuli were present, those on which non-target stimuli were present but did not move (static), and those on which non-target stimuli were present and did move (dynamic). When non-target stimuli were present they varied according to how many there were (4, 8 or 16), and, when they moved, they varied according to their velocity relative to the target (same, or +/- 50%), and their direction of movement (0-315 degrees in 45-deg, counter-clockwise increments)

Participants

A total of 44 Air Force recruits participated at the start of this study as part of their basic training requirements. All (but one) were right-handed, had normal or corrected to normal vision, and participated according to standard Air Force privacy and confidentiality procedures. Two different computer systems were used (see below) and five participants from each had their data deleted because the participants either did not understand or follow the instructions. These individuals were identified by having a very large number of repeated trials relative to the majority of participants. The final distribution included 8 males and 9 females having used a Dell® computer system, and 7 males and 8 females having used a Micron® computer system.

Materials and computers

The two-dimensional scenes were programmed to have a light gray background, black vertical start and finish lines positioned in the middle third of the screen, dark gray square targets, and somewhat lighter, square non-target stimuli (approximately 83-, 0-, 16-, and 39-% of "pure" white, respectively). So, the targets

were made darker to distinguish them from the stimuli. Brightness settings at all stations were equated by turning all monitors to the brightest level. This had an overall effect of reducing contrast, but still clearly retaining the distinction between target and non-target stimuli. In any condition, when the target and a non-target stimulus overlapped or intersected, the target appeared to be in front of the non-target stimulus. All paths "traveled" by the target had the same finish line, but the start lines varied (see Table 1), and all movements were from left to right.

Each scene came on and remained static (nothing moving) until the subject pressed the spacebar to initiate that trial. Initially, the target was entirely visible, its trailing edge at rest against the starting line. When the participant depressed the spacebar the visible target traveled for 2-sec before it disappeared.

The targets traveled at six different velocities (see Table 1 for specifications of distance, velocity and *TTC*), two different velocities and distances after disappearing for each of the three times to contact. The non-target stimuli traveled at one of three different velocities relative to the target depending on which condition the participant was in. One third of the participants saw the non-target stimuli moving at the same velocity as the target, one third saw them moving 50% faster than the target, and one third saw them moving 50% slower than the targets. On any given trial all the non-target stimuli moved in the same direction, and followed a path defined by degree of deviation from horizontal (in increments of 45-deg, counter-clockwise from the horizontal, left-to-right direction of 0-deg).

When they were present, there were either 4, 8 or 16 non-target stimuli, randomly positioned on the screen at the start of each trial. Initial non-target stimulus positions were determined by randomly choosing an x,y intersection from an imaginary 16x16 grid that filled nearly all of the viewable area on the computer monitor (inset about 2.54-cm on all sides), with the restriction than no x or y value was repeated. If and when amoving non-target stimulus "left" the screen, a new one immediately appeared and began to move at a location at the other end of an imaginary circular path around the screen. Figure 1 shows examples of scene presentations for 4, 8, and 16 stimuli, and also indicates examples of three of the eight different non-target stimulus movement directions.

The six-item *TTC* matrix (two different scene conditions for each of the three *TTC* durations, 2-, 4-, and 8-sec, as in Table 1) was crossed with the three levels of Number of non-target stimuli and the eight levels of Direction of movement, for a total of 144 trials. There were also two replications of the *TTC* matrix on which no

Table 1

Target and Stimulus Specifications

Overall distance (degrees of visual angle)	Velocity (degrees/second)	Time to Contact (seconds)
23.5	5.9	2.0
17.8	4.5	2.0
17.6	2.9	4.0
15.2	2.5	4.0
13.4	1.3	8.0
11.2	1.1	8.0

Length of “start” and “finish” vertical lines was 5.6 degrees of visual angle.

The target and distractor squares had sides of .7 degrees of visual angle.

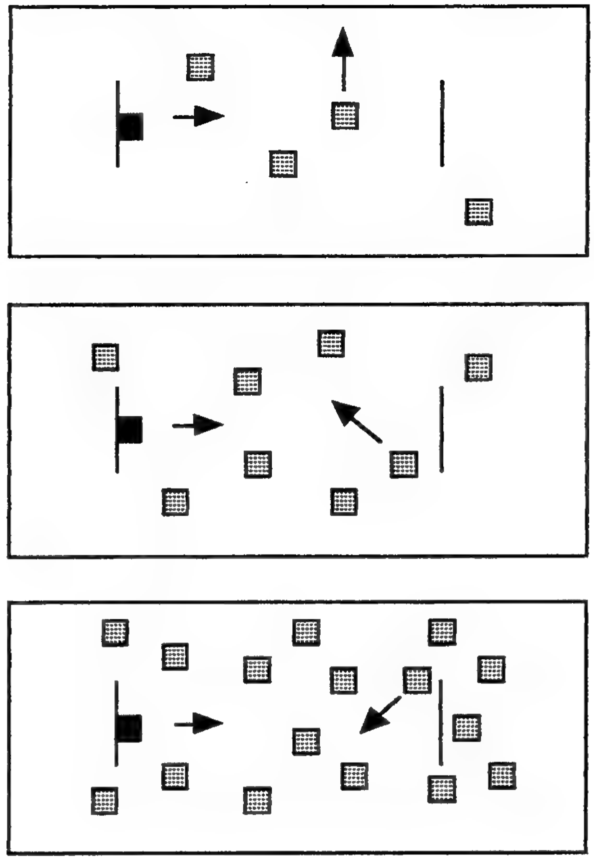


Figure 1. This figure shows samples of the three different numbers of non-target stimuli (4, 8, and 16), and three of the eight different directions of movement of the non-target stimuli.

non-target stimuli were present (12 trials), and six replications of the *TTC* matrix on which 2, 4, and 8 non-target stimuli were present (36 trials). Thus, there were a total of 192 unique trials.

In each session some participants used either a Dell® computer configured with a 90-MHz Pentium® processor with 16 megabyte of RAM, and a 17-in color monitor set to a black and white monochrome screen, or a Micron® computer using a 166-MHz Pentium® processor with 16 megabyte of RAM, and a 17-in color monitor set to a black and white monochrome screen The programs operated in EGA video, with a frame presentation rate of 14-msec per frame. We had no basis for predicting differences in performance based upon the systems, especially since frame rate was the same in both. In fact, a t-test on overall mean *TTC* estimates between the two systems resulted in an insignificant difference, 3.72-sec for the Dell® versus 3.81-sec for the Micron® [$t(30) = -.31$, $p > .75$], so the data from the two systems were pooled in the analyses presented below.

Procedure

The participants were run, on consecutive days, in two group sessions of 22 participants each. They were pseudo-randomly assigned to one of the computer stations, with the only restriction being that we attempted to evenly distribute men and women across computer systems and relative non-target stimulus speed conditions (normal, slower, and faster). The first part of the program described the use of the system, and demonstrated the stimulus conditions to be encountered during the study. There were also several practice trials with no non-target stimuli present, and which used a starting location longer than those used in the study, and a different (yet similar) velocity than any experienced in the study.

The presentation of criterion trials followed. To initiate each trial, the participant pressed the keyboard spacebar with left hand fingers to start the target moving, and pressed a mouse key using right hand fingers to make the *TTC* response. Upon the conclusion of the *TTC* response no feedback was given, and the scene for the next trial immediately appeared. The sequencing of the 192 trials was randomly determined for each participant. To compensate for inadvertent responses and possible inattention, a trial on which a *TTC* response occurred before the target had disappeared was aborted and was presented again at the end of the original series, as was any trial for which the *TTC* response was shorter than .5-sec. or longer than 12-sec. No trial was repeated more than once, and the average number of repeated trials was 13.35 (sd = 11.2), or just about 7%. Finally, participants proceeded at their own

pace with two one-minute rest breaks (remaining in their chairs and posturally oriented) after the 64th and 128th trials.

An important point needs to be introduced at this juncture. On day one of data collection there was an unplanned environmental occurrence, with the air conditioning in the testing center shutting down. Since the experimental sessions were conducted in mid-summer, the temperature and humidity in the testing center on that day became high enough to produce obvious general discomfort. Environmental data recorded in the testing center showed that the temperature had risen to 90°-F, with a humidity reading of 76%, sufficient to qualify for a "Category II" apparent heat index of approximately 110° F which can be associated with heat exhaustion in instances of prolonged physical activity (Steadman, 1979). Decrements in performance on visual processing tasks also have been found at this temperature (Hohnsbein, Peikarski, Kampmann & Noack, 1984). On day two of data collection the malfunction had been repaired, and readings were a much more comfortable 76°-F, with 72% humidity. In effect, we had an unplanned source of variance, a new factor - moderate heat-induced stress. This heat stress factor is introduced in the following analyses as the Day factor - high heat for day one, and normal conditions on day two.

Results

In each of the analyses that follow, mean *TTC* scores were computed over trials with actual *TTC* times of 2-, 4-, or 8-sec (respectively) in each condition, and those means were the data entered into the analyses of variance. Thus, for the no non-target stimuli and static non-target stimuli conditions, each *TTC* entry for each participant was based on four trials (observations), while in the dynamic non-target stimuli condition each *TTC* entry was based on two trials.

Does the presence or mere movement of non-target stimuli affect *TTC* accuracy? To answer this question *TTC* performance was assessed across the three task conditions with Gender and Day as between-subjects variables, and Task and *TTC* as within-subjects variables. That analysis yielded only an overall effect for *TTC*, $\underline{F}$ (2, 56) = 187.06, $\underline{p} < .0001$, with mean estimated *TTC* increasing as actual *TTC* increased, 2.17-, 3.68-, and 5.58-sec for actual *TTC* times of 2-, 4-, and 8-sec, respectively. No other main effects or interactions were significant at the .05 level. Thus, the mere presence (static condition) or movement (dynamic condition) of non-target stimuli did not have a significant affect on overall *TTC* estimates relative to the simple condition where only the target was present.

Does the number of non-target stimuli present have an effect on *TTC* accuracy? To answer this question an analysis of variance was performed on data from the static and dynamic conditions. In the latter, performance was pooled over

the direction of movement manipulations. This analysis had Gender and Day as between-subjects variables and Task, Number of non-target stimuli (4, 8, or 16), and *TTC* as within-subjects variables. There was a significant effect for *TTC*, $\underline{F}$ (2, 56) = 173.13, $\underline{p}$ < .0001, with increasing mean *TTC* estimates of 2.18-, 3.69-, and 5.55-sec. There was also a significant interaction between Day and Number of non-target stimuli, $\underline{F}$ (2, 56) = 3.85, $\underline{p}$ < .05. Day 1 (Heat) participants gave slightly longer estimates of *TTC* than Day 2 (Normal) participants, especially for the eight non-target stimuli condition. Although we had no a priori hypothesis about the effects of heat, it might be that the high heat and moderate numbers of non-target stimuli combined to create an optimum arousal-optimal performance situation, but such an explanation is purely speculative, and, in any event, Day (testing temperature) did not interact with *TTC* duration. The number of non-target stimuli yielded no main effect, nor did that factor interact with *TTC*.

Do the number, relative speed and direction of movement of non-target stimuli have an effect on *TTC* performance? To answer this question an analysis of variance was conducted on data only from the dynamic condition. That analysis had Gender, Day, and Relative Speed of non-target stimuli as between-subjects factors, and Number of non-target stimuli, Direction of Movement, and *TTC* as within-subjects variables. That analysis yielded a significant main effect for *TTC*, $\underline{F}$ (2, 40) = 137.25, $\underline{p}$< .0001, with increasing mean *TTC* scores of 2.12-, 3.66-, and 5.58-sec. There also was a significant interaction between Direction of movement and *TTC*, F (14, 280) = 4.83, $\underline{p}$< .0001, and between Relative Speed of movement of non-target stimuli, Direction of movement, and *TTC*, $\underline{F}$ (28, 280) = 1.91, $\underline{p}$ < .01. This latter interaction, encompassing the effects of the former, is shown in Figure 2. Time to contact estimates increased as a function of actual *TTC*, but there was the usual greater degree of underestimation of *TTC* as actual *TTC* increased. Further, with non-target stimuli moving at the same speed as the target, participants were much more accurate in their *TTC* estimates at the longest *TTC* duration when the non-target stimuli traveled in the same direction as the target. In this instance, underestimation was virtually eliminated. No other effects were significant at the .05 level.

What is the nature of individual differences in *TTC* performance? To answer this question we constructed an individual differences variable representing overall *TTC* performance in each of the three conditions. We chose the slope of the linear regression line fitted to each participant's mean judged *TTC* for the three actual *TTC* values of 2-, 4-, and 8-sec in each condition. The distributions of slope values for each task condition are shown in Figure 4. A slope value of 1.0 indicates perfect accuracy in *TTC* ability, while values less than 1.0 indicate a tendency towards underestimation,

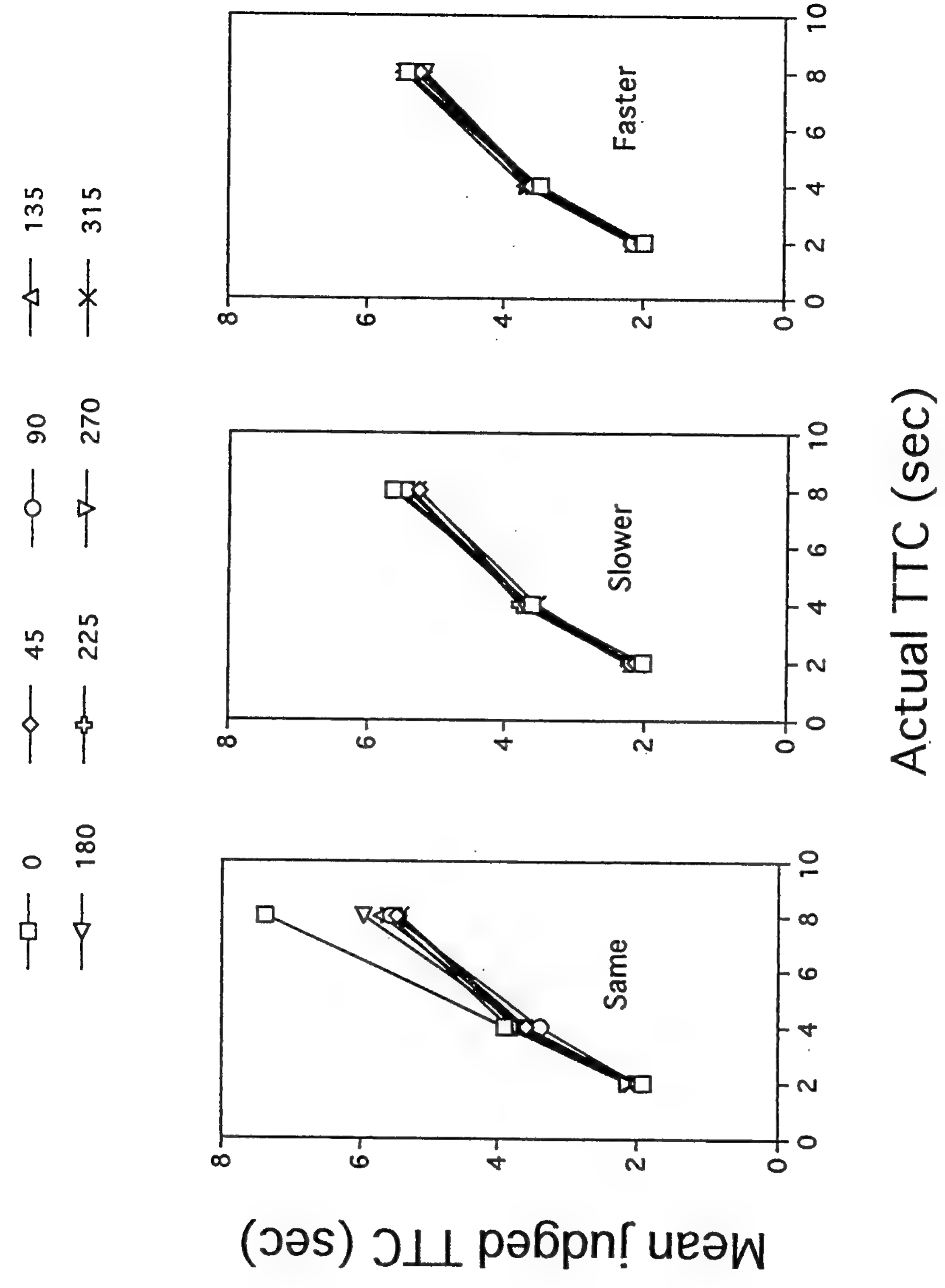


Figure 2. This figure shows the interaction between Speed and Direction of non-target stimuli and *TTC*.

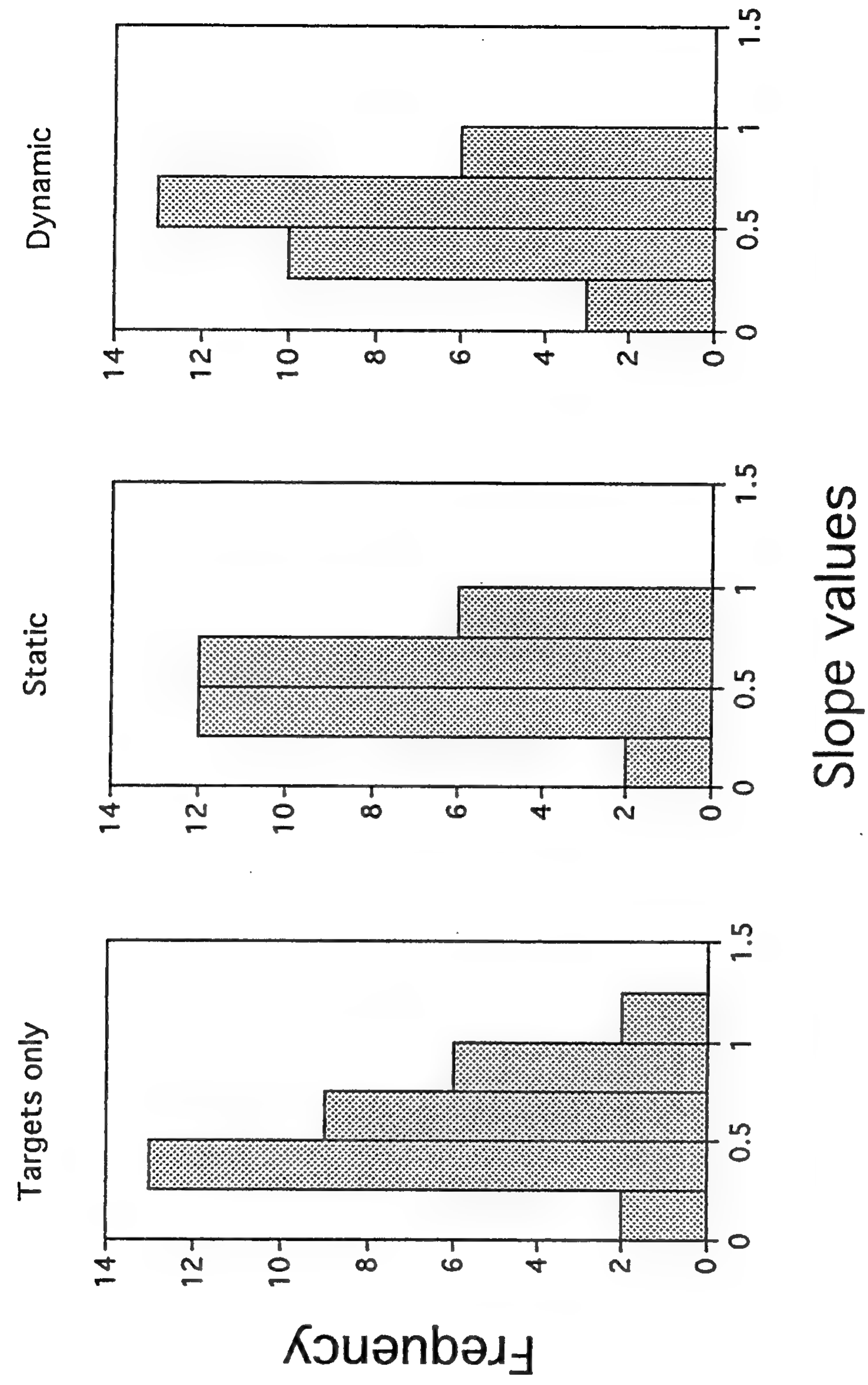


Figure 3. This figure shows the distribution of *TTC* slope functions under the three task conditions.

and values greater that 1.0 indicate a tendency toward overestimation of *TTC*. Nearly every slope value is less than 1.0, but one can observe a substantial range of values, and the possibility of an underlying normal distribution of *TTC* performance accuracy.

Discussion

We began this study with some expectation that the non-target stimulus manipulations would produce deleterious distraction effects on *TTC* performance, but we were not sure how those effects would be manifested in performance. It appears from our results, however, that *TTC* performance is rather difficult to disrupt. Non-target stimuli, even when they are numerous and moving across the target's path, do not seem to disrupt *TTC* judgments. We also had the opportunity to observe that not even a very hot and uncomfortable task environment produced a disruptive effect. In fact, the only substantial effect on *TTC* performance, other than the obvious effects of actual *TTC*, was the improvement in accuracy when the non-target stimuli moved at the same speed, and in the same direction as the target at the 8-sec *TTC*, but there is a plausible explanation for that facilitation effect.

A non-target stimulus traveling in the same direction and at the same speed as the target stimulus is essentially a running mate, and can become a surrogate target when the actual target disappears. One merely has to make a mental note of the degree of separation between the target and a correlated non-target stimulus, and use the movement and location of the surrogate non-target stimulus as a guide to when the target would reach the finish line. The longer the remaining travel time before the target would have contacted the finish line, the more time for the participant to make these mental compensations, and hence performance at the 8-sec *TTC* received most of the benefit of the surrogate process. Non-targets moving at different speeds or directions would serve the surrogate function less well, if at all, and that also is consistent with our findings. Unfortunately, we did not collect interview data following the task, so we have no direct confirmation of the surrogate process. This surrogate facilitation effect could be confirmed and further investigated in experimental situations in which, for example, the non-target stimuli themselves disappear sooner or later during the target's invisible period. The sooner they disappear, the less effective surrogates they would become.

Tentative acceptance of the surrogate explanation does confirm, somewhat, our initial speculation that the simple, target only, laboratory tasks could be overly simplified representations of conditions encountered in the real world. Apparently, in our much richer dynamic displays, our participants found a correlated (predictive) cue, the surrogate, to help them with the task. In fact, it may have

helped them so much that the usual increasing deviation from the actual *TTC* could be eliminated in some situations (e.g., the same velocity, 0-deg, 8-sec *TTC* condition). Further, real world situations are often replete with the presence of, and the opportunity to make use of, such "decision aids," so the surrogate effect should not be dismissed lightly. Rather, it should be recognized as being an instance of the use of general, strategic, and adaptive cognitive functions.

We considered that a motion repulsion effect might operate to influence the perception of the path of motion of the target, especially after it had disappeared from the screen. We found nothing to support motion repulsion effects in our data, and that is probably because the finish line was always visible to be a heading cue. In the absence of a visible finish line (one that disappears along with the target), or under conditions where the finish lines vary in direction, there may very well be a greater opportunity to observe *TTC* biases consistent with motion repulsion effects.

One of our original speculations was that non-target stimuli might consume some of the limited attentional resource available for the *TTC* task and decrease performance, but we have little to offer to confirm that notion. Attentional resources may not have been diverted by the non-target stimuli. Or, attentional resources may have been consumed by non-target stimulus conditions, but there may have been sufficient resources remaining to time share the *TTC* tasks. Or, *TTC* tasks may not require attentional resources to be performed. Hasher and Zacks (1979), among others, have suggested that, for humans, certain types of encoding operations require little or no attentional resources, and these have to do with the flow of information. While their concern was with memory operations encoding frequency, temporal order and spatial information, their general framework might extend to simple timing phenomena as well, since timing is essential in monitoring information flow. Indeed, there has even been speculation that the ability to time the release of projectiles, with the intent of hitting a stationary or moving object, might even be the basis for a hominid evolutionary drive (Calvin, 1983). Certainly, *TTC* estimation would be representative of such abilities, and it would serve evolutionary purposes well (not to mention individual survival!) if such abilities were not easily disrupted.

In the main, and under the limited conditions the present study, the results emphasize the robust nature of *TTC* decision operations. They are difficult to disrupt, and do not appear to be affected adversely by the presence of various numbers of non-target events, even when they move in a direction opposite to the target. Researching a related perceptual ability, Royden and Hildreth (1996) and others (Cutting, Vishton & Braren, 1995; Warren & Saunders, 1995) have made similar

observations with respect to research on heading judgments, concluding that "...moving objects do not significantly affect an observer's heading judgments in real situations..." (p. 851). Time-to-contact decisions may very well be based on similarly durable processes, but just how durable must be decided by further research.

References

Caird, J.K., & Hancock, P.A. (1994). The perception of arrival time for different oncoming vehicles at an intersection. Ecological Psychology, 6, 83-109.

Calvin, W.H. (1983). A stone's throw and its launch window: Timing precision and its imlications for language and hominid brain. Journal of Theoretical Biology, 3, 115-124.

Cutting, J.E., Vishton, P.M., & Braren, P.A. (1995). How we avoid collisions with stationary and moving objects. Psychological Review, 102, 627-651.

Hasher, L. & Zacks, R. (1979). Automatic and effortful processes in memory. Journal of Experimental Psychology: General, 108, 356-388

Lee, D.N. (1976). A theory of visual control of braking based on information about time-to-collision. Perception, 5, 437-459.

Marshak, W. & Sekuler, R. (1979). Mutual repulsion between moving visual targets. Science, 205, 1399-1401.

Royden, C.S. & Hildreth, E.C. (1996). Human heading judgments in the presence of moving objects. Perception & Psychophysics, 58, 836-856

Schiff, W. & Detwiler, M.L. (1979). Information used in judging impending collision. Perception, 8, 647-656.

Steadman, R.C. (1979). The assessment of sultriness. Journal of Applied Meteorology, 18, 861-884.

Tresilian, J.R. (1991). Empirical and theoretical issues in the perception of time to contact. Journal of Experimental Psychology: Human Perception and Performance, 17, 865-876.

Tresilian, J.R. (1995). Perceptual and cognitive processes in time-to-contact estimation. Analysis of prediction motion and relative judgment tasks. Perception and Psychophysics, 57, 231-245.

Warren, W.H. & Saunders, J.A. (1995). Perceiving heading in the presence of moving objects. Perception, 24, 315-331.

Acknowledgements

We thank the Air Force Office of Scientific Research for supporting this research, Dr. William Tirre of Armstrong Laboratory, Brooks Air Force Base, San Antonio, Texas for encouraging and facilitating all aspects of this project, and Ms. Karen Raouf, whose programming expertise was essential. We especially thank our families for cheerfully coping with our long absences during the conduct of this study. Comments may be sent by email to *p.h.marshall@ttu.edu.*

PERCEPTUAL ISSUES IN VIRTUAL ENVIRONMENTS AND OTHER SIMULATED DISPLAYS

Kelly G. Elliott
Graduate Student

School of Psychology
Georgia Institute of Technology
Atlanta, GA 30332-0170

Final Report for:
Graduate Summer Research Program
Armstrong Laboratory/CFHP

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Armstrong Laboratory/ CFHP

September, 1996

PERCEPTUAL ISSUES IN VIRTUAL ENVIRONMENTS AND OTHER SIMULATED DISPLAYS

Kelly G. Elliott
Graduate Research Assistant
School of Psychology
Georgia Institute of Technology

Abstract

Virtual environments are multisensory and highly interactive display systems that come in a myriad of flavors and varieties. These VE systems can serve a multitude of purposes within the scientific, medical, military, industrial, and entertainment fields in ways that more traditional human-computer interfaces simply cannot. Because of all the potential and actual uses for VE systems, developing an optimal VE system is a high priority.

Developing an optimal VE system requires knowing and capitalizing on the capabilities and limitations of human perception, both within a given sensory modality and integrated across sensory modalities. Yet, no available VE system can fully exploit the capabilities of human perception, especially those of human vision. These technological limitations can impose some perceptual tradeoffs in utilizing available VE systems that one must carefully consider.

Conversely, VE systems provide an opportunity to answer some fundamental questions about how humans build up percepts about what is out there and what is going on, both within a given sensory modality (e.g., vision) and integrated across sensory modalities. Although VE systems ideally could mimic real-world experiences, they are not bound by the limitations of the real world (e.g., gravity and the laws of physics). Thus, perception in the simulated world of a VE system can dramatically differ from that in the real world. That is, VE systems allow us to test some limitations and capabilities of human perception in ways that more traditional displays and the real environment do not.

Our challenges this summer were to tackle these issues by thinking and reading, by setting up and conducting some pilot studies to explore the formation of multistable percepts within a virtual environment, and by writing a draft of a review paper based on these ruminations and preliminary results.

PERCEPTUAL ISSUES IN VIRTUAL ENVIRONMENTS AND OTHER SIMULATED DISPLAYS

There has been alot of hype in the media about virtual environments and cyberspace. Although there are innumerable definitions of a virtual environment, Kalawsky (1993) defines it as a synthetic sensory experience that communicates physical and abstract components to a human user or participant. The basic components of a VE system consist of a system for generating the virtual environment, a visual display, an auditory display, possibly a tactile display, a system for tracking head and hand as well as possibly tracking eye or body positions, a manipulandum (e.g., a dataglove or 3-D joystick) and/or a speech recognition interface for communicating with the virtual environment. But, the actual components of the VE system may vary, as may the specific implementations of each component.

A myriad of actual and of potential uses exist for virtual environments -- within scientific and medical areas, within both the military and industrial arenas, and for use in the field of entertainment. Virtual environments can be used for training, as in flight training and simulation, driver training and assessment, surgical training, and training astronauts to deal with nonterrestrial environmental conditions. They can aid medical doctors in actual surgical operations and scientists in visualizing spatial relationships and how different parts interact, as in the visualization of planets or the molecular dockings of atoms. Virtual environments can be used in the teleoperation of robots through hazardous or unknown terrains -- as in locating and removing a bomb in a building or in exploring Jupiter's surface. Architects can walk their clients through a virtual building, so that costly modifications and adjustments can be avoided. Virtual environments can be used for sheer entertainment, too. For instance, one could take a virtual vacation to the location of one's dreams sans air travel or explore the ruins of Mesa Verde as they appeared to inhabitants almost a thousand years ago. These virtual environment pursuits places various demands both on the hardware system and the human perceptual and cognitive system that processes these synthetic inputs.

How do we actually build up our percepts of what is out there and of what is going on in either a real or a virtual environment? Although most virtual environments emphasize visual input, experiences within a virtual environment usually are multisensory and highly interactive -- as they are in real environments. Yet, perception in the simulated world of a virtual environment can differ from real

world experiences. Some of these differences are intentional and some are coincidental. What are some of the perceptual tradeoffs and limitations when interacting with a virtual environment? What future challenges face the designers and developers of virtual environments?

These are questions we posed and began investigating this summer. The results of our ten week tenure at Wright Patterson Air Force Base include thinking and reading about these issues, setting up and conducting some pilot studies to explore the formation of percepts within virtual environments, as well as drafting a review paper for publication based on some of these ruminations and preliminary results.

How perception in the simulated world of VE differs from perception in the real world

Virtual environments often try to mimic the real-world environment. Yet, the very nature of virtual environments suggests we can simulate information not readily available to our senses in real-world environments and that we can interact with this virtual environment in ways that normally are prevented by real-world constraints. For example, a VE could allow us to defy gravity by floating through space (as in microgravity conditions) or walking on ceilings. A VE could also visually (or auditorally) present distance information from a laser range finder. It could present infrared or thermal images to aid the identification of objects viewed under low ambient luminance conditions. In short, a VE system can overcome some of the limitations of the human perceptual system and add to the store of information normally available.

A VE system also can eliminate perceptual information normally available to our senses in the real world (e.g., eliminating the 3D visual information normally provided by stereopsis or the kinesthetic feedback from touching and pushing against objects). Sometimes information may be eliminated on purpose because the additional information is not *cost-effective* and/or adds no useful information for the tasks performed. That is, sometimes information is eliminated or included by purposeful design to enhance perception in and interaction with virtual environments. But, sometimes information is eliminated because of current technological limitations in the available hardware and software.

In any case, there are always tradeoffs -- enhancing one type of information or capability may adversely affect the presentation and availability of other

information and capabilities. On one hand, accessing more information can overload the human or machine system (or both!) thus causing information processing to become resource limited and non-optimal (Norman & Bobrow, 1975; Stokes, Wickens, & Kite, 1990). On the other hand, severely degraded displays that greatly restrict available information can cause data-limited deficits in information processing that also are not optimal.

Some perceptual tradeoffs and limitations of perceiving and of interacting within a simulated world

Although virtual environment systems come in many flavors and varieties and have countless potential uses, all VE systems face challenges in their design, development, and implementation. For example, all VE systems currently lack the full fidelity of real-world experiences and many of them do not incorporate certain sensory inputs (e.g., haptic, tactile and force feedback as well as taste or smell).[1] Some of the challenges faced are similar to those encountered with more traditional HCI systems -- including flight simulators -- and, thus, VE design and development can benefit from some of the previous lessons learned.

One top priority is meshing the human-computer interface design of a VE system so that it exploits the capabilities and limitations of the human perceptual system as well as those of the virtual environment. Moreover, designers and developers are working on incorporating some of the missing features into VE systems, but hopefully only when they are deemed cost-effective and can enhance actual task performance. Although several pressing problems face designers and developers of virtual environment systems, not even the users of virtual environments agree which problem is the most pressing and critical. Several candidate perceptual problems in vision and audition are discussed here.

A. Some visual considerations

Some feel the most challenging problems within VE systems may be visual, given that humans depend so heavily on visual information about their real-world environments and that at least 50% of the human brain is involved in processing visual information.

---

[1]Sensorama is an early virtual environment system that actually did include olfactory information from various smells as the user traversed the virtual roadway (Kalawsky, 1993; Rheingold, 1991).

Field of view (FOV) vs. spatial resolution tradeoff. Of these visual problems in VE, one of the most pressing is to simultaneously provide the user with a wide field of view *and* with good spatial resolution. No existing virtual environment yet achieves the wide field of view of the human visual system operating in a real environment: The human's instantaneous binocular field of view is approximately 200 deg in the horizontal direction and 120 degrees in the vertical direction. But, to create an immersive visual environment, designers want a very wide field of view (wFOV). Unfortunately, such a wide field of view often leads to worse spatial resolution within the visual image, so that spatial details of objects are lost and curved or diagonal edges appear jagged (i.e., spatial aliasing occurs). This is the well-known FOV versus spatial resolution tradeoff faced by designers of visual displays. Although humans can easily discern spatial details of even less than 1′ of visual angle (corresponding to a Snellen visual acuity of 20/20 or better), most virtual environment displays cannot provide such fine spatial resolution.

Some potential solutions to the FOV versus spatial resolution tradeoff problem include: (a) tiling multiple displays (e.g., the multiple LCS of Kaiser Electro-optics HMDs); (b) using a mixed spatial resolution to provide a higher spatial resolution within the central 2 degrees to 5 degrees of the visual field (where human visual acuity is best) but lower spatial resolution outside this region (e.g., CAE Electronics' high-priced fiber-optic HMD); (c) use binocular or biocular displays with partial overlap of the visual fields instead of complete overlap; (d) using multiple projections onto surrounding walls or screens (e.g., the CAVE or the Virtual Workbench); or (e) designing a new and better display element.

Each of these potential solutions has additional technological problems and each can significantly increase the complexity and cost of the display. For example, a mixed resolution display assumes *line of sight* viewing -- where the user *always* looks straight ahead rather than off the to side. Moreover, the sense of immersion created by a wider field of view does not necessarily result in better performance or greater user comfort.

Stereoscopic displays vs. biocular displays. Others feel that the 3D stereoscopic imagery created with binocular displays can provide a compelling sense of presence and that honing stereoscopic displays is the most pressing VE challenge. Stereopsis certainly can enhance performance of close work involving fine motor control, such as surgical operations, as well as long-range distance perception of an object one mile away from another at optical infinity. In a binocular display a slightly different image is presented to the right eye from that presented to the left

eye, unlike a biocular display in which exactly the same image is presented to both eyes. Stereopsis results from the fusion of the two slightly different views that the user's laterally-displaced eyes receive in viewing a real or a simulated 3D stereoscopic display (Arditi, 1986; Davis & Hodges, 1995; Tyler, 1983).

Because stereoscopic depth depends on the interpupillary distance (IPD) between the two eyes, the optics of the binocular display or the binocular parallax of the computer graphics and/or the offset of the two visual sensor devices should be adjustable. The average user's IPD is approximately 63 mm, with a range of 53 mm to 73 mm (Kalawsky, 1993). The larger the IPD, the more salient the perceived stereoscopic depth -- so artificially increasing the *virtual* IPD could enhance the perceived stereoscopic depth. But, beware! Creating too much binocular parallax can destroy perceptual fusion, resulting in diplopia or double-vision with accompanying feelings of nausea, eye strain, and possible performance deficits (Piantanida, Boman, & Gille, 1995).

Using a binocular (or biocular) visual display with only partial overlap can result in a wider FOV, while maintaining an adequate spatial resolution, as previously explained. For most humans, in viewing the real world, the region of binocular overlap is 120^{o}, with a monocular visual field of approximately 35^{o} flanking each side of the binocular overlap. For most VE systems with partial overlap, however, the region of overlap is considerably less than 120^{o}. Also, with partial overlap there are flanking right and left monocular regions of the display; each monocular region provides very different spatial and luminance configurations for the right versus the left eye. Because of these two factors, the user perceives binocular rivalry and monocular suppression within each flanking monocular region of the display, with concomitant discomfort and disorientation. Potential solutions to these problems include (1) a relatively large region of overlap, one that more closely approximates the 120^{o} binocular field of natural viewing; (2) blurring the edges of the display so that the entire visual field appears less fragmented (Melzer & Moffitt, 1989; Melzer & Moffitt, 1991); and (3) using *convergent* rather than *divergent* binocular displays (Melzer & Moffitt, 1989; Melzer & Moffitt, 1991). As a *rule of thumb* Kalawsky (1993) recommends using complete binocular overlap when moderate FOVs suffice, but partial overlap only if very wide FOVs are necessary.

Binocular displays are heavier, more complex, and costlier than are simpler biocular displays, although both require adjustable IPDs and careful alignment of each eye's view. Still, not everyone possesses good stereopsis so binocular displays

have limited effectiveness for these individuals. There are amblyopes who have no stereopsis as well as some individuals who are stereoanomalous. Richards (1971) has shown that stereoanomalous observers may be able to perceive stereoscopic depth for objects presented in front of the point of fixation, but not behind it (or vice versa). Even some individuals with normal stereopsis may have difficulty perceiving stereoscopic depth in a visual display, although with training and feedback their stereoscopic visual performance sometimes improves dramatically (McWhorter, 1993; Surdick, Davis, King, & Hodges, 1997).

So, given the above constraints, when are binocular VEs more useful than biocular ones? Binocular VEs are useful when a visual scene is presented in a perspective view than in a bird's eye view (Barfield & Rosenberg, 1992; Yeh & Silverstein, 1992), when monocular cues provided by a biocular display provide ambiguous or less effective information than the stereopsis provided by binocular HMDs (Yeh & Silverstein, 1992), when static or slowly-changing visual displays are used rather than rapidly-changing, dynamic displays (Wickens, 1990; Wickens & Todd, 1990; Yeh & Silverstein, 1992), when ambiguous objects or complex scenes are presented (Cole, Merritt, & Lester, 1990; Spain & Holzhausen, 1991), and when complex 3D manipulation tasks require ballistic movements or very accurate placement and positioning of objects or tools. Stereopsis is helpful in these situations for a two primary reasons. First, it helps disambiguate elevation and distance information in providing information about the spatial layout of objects (e.g., in a perspective view). Second, it provides the user with fine depth discrimination for objects (and the shapes of objects) located within an arm's length of the user.

Color versus monochrome. The issue of color displays in virtual environments is whether one should use color or not. The use of color within the display can add to a sense of immersion and realism. Users often prefer color displays and color-coding within a display can be helpful, especially in virtual environments where both the rate and amount of information transmitted is high (Christ, 1975; Stokes, et al., 1990). In short, some reasons color displays are desirable include: (1) color can be used to unify or cluster disparate elements of a display (Christ, 1975; Christ & Corso, 1982); (2) color seems to be processed earlier and faster than other types of visual information, such as shape; (3) objects are more easily identified by color than by other features based on size, shape, or brightness; (4) color coding can significantly reduce visual search time; and (5) color adds to the *realism* of the virtual environment.

Given all of these advantages, why would one consider *not* using a color display within a virtual environment? The reasons not to use color primarily involve technological limitations and cost factors. With most existing displays the use of color involves tradeoffs with other capabilities of the display such as spatial resolution or temporal resolution. Available color displays lack the full range of chromaticities (hues and saturations) that the human can perceive. They also tend to have less contrast or brightness than monochromatic or grayscale displays as well as being heavier and costlier. Although approximately 8% of the male adult population have inherited color vision deficiencies that limit the usefulness of color displays[2], most adults (both male and female) will *acquire* color vision deficiencies as a result of normal aging processes or of disease processes (e.g., diabetes). Moreover, in low luminance environments, such as those simulating nighttime vision, all users are blind to color -- instead they see only in black and white and shades of gray.

Temporal resolution, update rates, and time lags. Still other VE developers and users feel that time lags caused by relatively slow update rates or frame rates are the most pressing challenge. Both frame rate and update rate affect the temporal resolution of a visual display. Frame rate is a hardware-controlled variable that determines how many images each eye sees per second (measured in Hertz), whereas update rate is the rate at which changes in the image are updated. Of the two, update rate is the more problematic. The update rate depends on the computational complexity of an image[3] and can be no faster than the frame rate. The update rate constrains how fast virtual objects can move around in virtual space yet avoid jerky movement.[4] Update rate also constrains how quickly the user can move his or her head yet have the virtual images remain meshed with the head movement.[5] If the temporal resolution of the visual display is too low, it can hinder task performance or cause illusory motion artifacts. In fact, noticeable time lags in update rates can cause nausea and disorientation in space -- making the user

[2]Only about 0.5% of females have inherited color deficiencies.

[3] Update rates also are influenced by time lags and delays in sensory devices (e.g., headtracking devices). The effects of computational complexity of images and of sensory delays are additive.

[4] The minimum update rate in Hertz is the virtual object's angular speed in arcmin/sec divided by 15, assuming a maximum displacement of 15' per frame for the perception of smooth movement (e.g., Braddick, 1974).

[5]When the user's head position is monitored and the visual scene is updated according to the head position, then update rate depends on sensor lag in determining the head position as well as on the computational complexity of the visual imagery. These two factors are additive.

very uncomfortable and, perhaps, hindering his or her performance (Piantanida, et al., 1995).

Meshing real and virtual imagery. In augmented-reality all or part of a real-world scene is combined with synthetic imagery. There are two approaches to creating this augmented reality. The most popular augmented reality system is to use a see-through head-mounted display in which computer graphics or symbology is superimposed on a direct view of the real world. An optical combiner composed of half-silvered mirrors is used to mesh the synthetic imagery with real-world scenes (Barfield, Rosenberg, Han, & Furness, 1992; Davis, 1996). Because the older heads-up display (HUD) is an example of this type of augmented reality, alot already learned about presenting symbology in a HUD is relevant here.

Another augmented reality system is to use an opaque HMD or projection screen system in which computer graphics is superimposed on a local or remote video of real-world scenes. Video keying is used to electronically merge video and synthetic images, similar to that used for CNN's news broadcasts. Computer graphics or other video clips are used to fill-in a large blue screen behind the anchor. The particular blue hue is used to mark the area for video keying because most human skin tones do not contain this color (Barfield, Rosenberg, & Lotens, 1995). (However, if the anchor wore a dress or jacket of that same blue color, TV viewers would have an eerie view of a vanishing anchorperson!)

Both methods of visually presenting augmented reality have some similar requirements. For both, accurate registration of the synthetic imagery and the real world is necessary. For example, meshing of perceived distance and of spatial layout matters. Roscoe (1984; 1993) has reported that virtual objects often appear farther away than do real-world objects; he suggests using a magnification factor of approximately 1.25 to correct for this disparity in the perceived distances of virtual and real objects. Although some have confirmed Roscoe's findings in an augmented reality system (Rolland, Gibson, & Ariely, 1995; Rolland, Holloway, & Fuchs, 1994), others have reported that in see-through HMDs virtual objects almost always appear closer (e.g., K. Moffitt, personal communication). Perhaps virtual objects appear closer in a see-through HMD because virtual objects occlude real-world objects or because accommodation cues signal that the virtual object presented on a visor is closer than the real objects viewed through the visor. In fact, when spatial layouts and updates of virtual imagery does not mesh with that of real-world scenes, this may introduce a new source of *simulator sickness* which includes vertigo, dizziness, and disorientation. It is also true that directly viewed real-world

scenes in a see-through HMD will mesh with vestibular cues resulting from turning one's head, et cetera; thus augmented-reality displays with see-through HMDs may reduce some of the simulator sickness symptoms reported with virtual environments created with opaque HMDs and updates rates that are "not fast enough".

For see-through HMDs there are some additional problems because the real-world scene effectively provides a background against which the synthesized images and symbology are viewed. At nighttime and under dim illumination, the effective background luminance provided by the real environment can be quite low, increasing the effective contrast and visibility of light symbols, but decreasing those of dark symbols. On a bright, sunny day, however, the effective background luminance provided by the real environment can be quite high, drastically decreasing the contrast and visibility of even the brightest synthetic symbology. To offset this problem one could change the overall gain of the real-world luminances to some constant value and, perhaps, simultaneously change the mean luminance and contrast gains in the synthetic imagery. Some lessons learned from using heads-up displays (HUDs) are relevant here -- the military currently uses a dark visor over the headset combiner to overcome the problem of a bright, ambient luminance from the real-world scene. Another possible approach would be to use light-sensitive filters over the headset combiner -- similar to that used in modern, high-tech sunglasses. Yet another possibility is the ability to change the contrast and polarity (i.e., light versus dark) of the virtual symbology.

We have emphasized visual augmented reality systems, which are the most highly developed augmented reality systems to date. Of course, augmented reality systems using other sensory modalities, such as auditory and tactile, are possibilities and also are worth exploring. After all, virtual environments are a multisensory and interactive experience. In the everyday real world individuals are rarely confronted with only visual information, but instead are simultaneously bombarded with a wealth of information from a variety of sensory modalities, including auditory information.

B. Some auditory considerations

While perception of 3D objects in VEs can be created using a wide variety of visual information, auditory inputs may be synchronized with visual inputs to provide an even greater sense of immersion and realism than that achieved by a single sensory modality (Bryson, Pausch, Robinett, &

van Dam, 1993; Burgess & Verlinden, 1993; Kalawsky, 1993). Although VE development has focused more on visual aspects of the environment than on other potential components of the system, with the introduction of more sophisticated and cost-effective audio hardware and system software it is now possible and prudent to use auditory information to enhance the realism and sense of presence within virtual worlds (Astheimer, 1993; Barfield & Furness, 1995; Bargar, Blattner, Kramer, Smith, & Wenzel, 1993).

Spatialized vs. non-spatialized (stereo) sound. Whether to include 3D spatial audio or simple stereo sound in simulated environments is an important issue for incorporating auditory displays into VE systems. Whereas simple stereo sound elicits a perceived position of the sound source along a one-dimensional line (i.e., the sound is heard as coming from the left or the right of the listener), spatialized sound is perceived as coming from a precise location in space and has not only a left-right attribute but also up-down and distance components (Barfield & Furness, 1995; Bryson, et al., 1993; Burgess, 1992).

Spatialized or true 3D sound was first created using a digital-signal processing device known as the Convolvotron, which was designed by Crystal River Engineering and NASA Ames Research Center (Bryson, et al., 1993; Burgess, 1992). This commercialized system transforms digital auditory signals in a similar manner to how the shape of the human's outer ear (or pinna) transforms real-world sounds (Barfield & Furness, 1995). That is, the Convolvotron uses head-related transfer functions (HRTFs) to filter the digitized sounds (Astheimer, 1993; Burgess, 1992). These HRTFs not only take into account pinnae effects but also interaural time differences (ITDs), interaural intensity differences (IIDs), and diffraction effects caused by the presence of the listener's head and body (Barfield & Furness, 1995; Durlach, 1991; Kalawsky, 1993). When the filtered sounds are heard by a listener, they appear to have a true sense of 3D location in space. Not only do these sounds have a directional component, but they also retain the acoustical properties of the listening environment (e.g., room reverberation cues) (Bryson, et al., 1993).

Spatialized sound is capable of providing an incredible sense of realism and presence in a VE. Furthermore, because of its spatial characteristics, it can be used to assign qualities to virtual objects. Spatial 3D audio is also very useful because it can provide crucial directional cues that better allow us to

navigate within the virtual world. Another advantage that spatialized sound has over simple stereo sound is that when the former is presented over headphones, for example within a HMD, it is perceived as coming from a particular location in space outside of the listener's head (Burgess, 1992). Simple stereo sound, however, is often perceived as originating from within the user's head (Barfield & Furness, 1995; Burgess, 1992). This perceptual problem, known as lack of externalization, is the direct result of stereo recording's poor model of how the human's pinnae alter sound stimuli once it arrives at the ear (Burgess, 1992).

Although spatialized sound can greatly enhance the spatial characteristics of a virtual auditory environment, it does have its drawbacks. In order to create spatialized sound, the acoustical properties of a particular listening environment must be modeled so that reverberation cues can be provided (Burgess, 1992; Burgess & Verlinden, 1993). Once a model has been developed, it must then be updated in real time. An even bigger drawback for spatialized sound is that with the Convolvotron's fast and complex computing power also comes its high cost. Incredible hardware costs have made this special processing device impractical for many VE system designers (Burgess, 1992). Despite the fact that some designers may choose to invest in a lower cost system, such as the presently-available Focal Point system, many system designers still choose to use simple stereo sound. Better, more affordable audio hardware and software is necessary before spatialized sound will be widely used by the VE community.

Real-world vs. synthesized sound effects. Must sounds used in VEs resemble ordinary real-world sounds or can artificially-created, synthetic sounds relay the same intuitive information in virtual worlds? The purpose of simulated environments, such as VEs, is not to create worlds which totally emulate the real world; however, system developers should be able to design realistic worlds if that is indeed the goal (Bargar, et al., 1993). Therefore, we probably want the capability of including both real-world and synthesized sound effects, depending upon the specific task employed within the VE and the precise goal of the environmental design.

Studies have revealed that users are more likely to respond favorably to real-world sounds than to artificial ones (Bargar, et al., 1993). In fact, natural sounds are often considered an effective type of auditory input in simulated environments merely because they are immediately recognizable by the

listener and therefore require neither adaptation nor training (Durlach, 1991). Although synthesized sounds may not be familiar to the listener, they can provide a greater range of distinct sound combinations than can real-world sounds (Bryson, et al., 1993). However, presenting virtual sounds that relate to the listener's everyday world experiences and that can be used in intuitive, self-evident ways is probably the most efficient and effective use of auditory input in simulated auditory display systems.

Meshing real and virtual auditory input. As previously mentioned, augmented reality displays do not have to be specifically visual in nature. In fact, auditory augmented reality systems can be used to enhance the real world in novel ways that allow the human to more efficiently and effectively interact with complex systems (Barfield & Furness, 1995). For example, digitized spatial audio can be overlaid onto a real sound in order to amplify the significance of the real-world noise. Conversely, anti-noise filters can dampen predictable or regular real-world noise.

C. Meshing multisensory inputs and Interactive Displays

Just as real and virtual imagery within a given sensory modality must mesh, so too must the different sensory inputs mesh together so that a coherent percept of the virtual world results. This is especially true for interactive displays in which the user moves around within the virtual environment and manipulates virtual objects. Time lags between inputs to the different sensory modalities can be confusing, disorienting, and downright nauseating. Not only must the temporal properties mesh across the different sensory modalities, but also the spatial properties must properly mesh together. Accurately perceived locations of objects in the azimuth, elevation, and distance (x, y, z) dimensions are important for the perception of where the objects are located and of the spatial layout of scenes or configurations. What happens when information provided by the different sensory modalities conflicts or is not in register across the senses? If visual and auditory information do not mesh together to form a coherent percept, often the visual sense will dominate -- a phenomenon known as visual capture (Welch & Warren, 1986). Although vision is our keenest sense, when it conflicts with tactile and kinesthetic inputs, it is the latter which may dominate (Welch & Warren, 1986).

Exactly how information is integrated within a given sensory modality or across different sensory modalities remains an unsolved puzzle. The solution to this puzzle can help to determine the relative weight of each sensory input in forming a coherent percept of the real or virtual world. It also can help us

determine how much error the system can tolerate in meshing together various sensory inputs.

Interacting with the environment sometimes can compensate for poor sensory or perceptual information within a virtual environment. For example, a study by Smets and Overbreke (1995) found that yoking a camera's view of a visual scene to the user's head movements improved the user's ability to solve a visual puzzle, despite relatively poor spatial resolution in the visual image. Results such as these suggest an ecological or Gibsonnian (Gibson, 1979) approach to the development and use of virtual environments: Performance within a virtual environment partially depends on the interactions between the perceptual quality of the displays and active manipulation of the environment.

In general, the more capabilities the virtual environment has, the more difficult it is to effectively interface this system with the human user. Yet, the more capabilities the virtual environment has, especially if these capabilities are calibrated with each other and mesh together, the more compelling the percepts formed of the objects, spatial layouts, scenes and configurations in the virtual environment.

References

Arditi, A. (1986). Binocular vision. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), Handbook of Perception and Human Performance, Vol. 1 New York: Wiley.

Astheimer, P. (1993). What you see is what you hear: Acoustics applied in virtual worlds. In Proceedings of the IEEE 1993 Conference.

Barfield, W., & Furness, T. A. (Eds.). (1995). Virtual Environments and Advanced Interface Design. New York: Oxford University Press.

Barfield, W., & Rosenberg, C. (1992). Comparison of stereoscopic and perspective display formats for exocentric judgment tasks. In Paper presented at the Annual Meeting of the Human Factors Society. Atlanta, Georgia.

Barfield, W., Rosenberg, C., Han, S.-H., & Furness, T. (1992). A god's eye (exocentric) versus a pilot's eye (egocentric) frame-of-reference for enhanced siuational awareness. In Proceedings of the Annual Meeting of the Human Factors Society.

Atlanta, Georgia: Working Paper: Department of Industrial Engineering, University of Washington.

Barfield, W., Rosenberg, C., & Lotens, W. A. (1995). Augmented-reality displays. In W. Barfield & T. A. Furness (Eds.), Virtual Environments and Advanced Interface Design New York: Oxford University Press.

Bargar, R., Blattner, M., Kramer, G., Smith, J., & Wenzel, E. (1993). Effective nonspeech audio for virtual reality. In Proceedings of the IEEE 1993 Conference.

Bryson, S., Pausch, R., Robinett, W., & van Dam, A. (1993). Implementing virtual reality. In Proceedings of the ACM SIGGRAPH 1993 Conference. Anaheim, CA.

Burgess, D. A. (1992). Techniques for low cost spatial audio. In Proceedings of the Fifth Annual Symposium on User Interface Software and Technology, . New York, NY: ACM.

Burgess, D. A., & Verlinden, J. C. (1993). An architecture for spatial audio server No. GIT-GVU-93-91). Graphics, Visualization, and Usability Center, Georgia Institute of Technology.

Christ, R. E. (1975). Review and analysis of color coding research for visual displays. Human Factors, 17, 542-570.

Christ, R. E., & Corso, G. (1982). The effects of extended practice on the evaluation of visual display code. Human Factors, 25, 71-84.

Cole, R. E., Merritt, J. O., & Lester, P. (1990). Remote manipulator tasks impossible without stereo TV. In J. O. Merritt & S. S. Fisher (Eds.), Proceedings of the SPIE: Stereoscopic Displays and Applications I, 1256 (pp. 255-265). Bellingham, WA: International Society for Optical Engineering.

Davis, E. T. (1996). Visual requirements in HMDs: What can we see and what do we need to see? In J. Melzer & K. Moffitt (Eds.), Head Mounted Displays: Designing VR System for the User New York: McGraw-Hill.

Davis, E. T., & Hodges, L. F. (1995). Human stereopsis, fusion, and stereoscopic virtual environments. In W. Barfield & T. Furness (Eds.), Virtual Environments and Advanced Interface Design New York: Oxford.

Durlach, N. (1991). Auditory localization in teleoperator and virtual environment systems: Ideas, issues, and problems. Perception, 20, 543-554.

Gibson, J. J. (1979). The ecological approach to visual perception. Boston: Houghton Mifflin.

Kalawsky, R. S. (1993). The Science of Virtual Reality and Virtual Environments. Wokingham, England: Addison-Wesley Publishing Company.

Melzer, J. E., & Moffitt, K. (1989). Partial binocular-overlap in helmet-mounted displays. In SPIE Proceedings: Display System Optics II, 1117 (pp. 56-62). San José, California: SPIE -- The International Society for Optical Engineering.

Melzer, J. E., & Moffitt, K. (1991). An ecological approach to partial binocular overlap. In S.P.I.E. Proceedings: Large Screen Projection, Avionics, and Helmet-mounted Displays, 1456 (pp. 175-191). San José, California: SPIE -- The International Society for Optical Engineering.

Norman, D., & Bobrow, D. (1975). On data-limited and resource-limited processing. Journal of Cognitive Psychology, 7, 44-60.

Piantanida, T., Boman, D. K., & Gille, J. (1995). Perceptual issues in virtual reality. Unpublished manuscript.

Richards, W. (1971). Anomalous stereoscopic depth perception. Journal of the Optical Society of America, 61(3), 410-414.

Rolland, J. P., Gibson, W., & Ariely, D. (1995). Towards quantifying depth and size perception in virtual environments. Presence, 4(1), 24-49.

Rolland, J. P., Holloway, R. L., & Fuchs, H. (1994). A comparison of optical and video see-through head-mounted displays. In S.P.I.E., 2351 (pp. 293-307).

Roscoe, S. N. (1984). Judgments of size and distance in imaging displays. Human Factors, 26(6), 617-629.

Roscoe, S. N. (1993). The eyes prefer real images. In S. R. Ellis, M. K. Kaiser, & A. J. Grunwald (Eds.), Pictorial Communication Virtual and Real Environments Washington, D.C.: Taylor & Francis.

Smets, G. J. F., & Overbeeke, K. J. (1995). Trade-off between resolution and interactivity in spatial task performance. IEEE Computer Graphics and Applications, September, 46-51.

Spain, E. H., & Holzhausen, K. P. (1991). Stereoscopic versus orthogonal view displays for performance of a remote manipulation task. In J. O. Merritt & S. S. Fisher (Eds.), Proceedings of the SPIE: Stereoscopic Displays and Applications II, 1457 (pp. 103-110). Bellingham, WA: International Society for Optical Engineering.

Stokes, A., Wickens, C., & Kite, K. (1990). Display Technology: Human Factors Concepts. Warrendale, PA: Society of Automotive Engineers, Inc.

Surdick, R. T., Davis, E. T., King, R. A., & Hodges, L. F. (1997). The perception of distance in simulated displays -- A comparison of the effectiveness and accuracy of multiple depth cues across viewing distances. Presence.

Tyler, C. W. (1983). Sensory processing of binocular disparity. In C. M. Schor & K. J. Ciuffreda (Eds.), Vergence Eye Movements: Basic and Clinical Aspects Boston: Buttersworth.

Welch, R. B., & Warren, D. H. (1986). Intersensory interactions. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), Handbook of Perception and Human Performance New York: Wiley.

Wickens, C. D. (1990). Three-dimensional stereoscopic display implemention: Guidelines derived from human visual capabilities. In J. O. Merritt & S. S. Fisher (Eds.), Proceedings of the SPIE: Stereoscopic Displays and Applications I, 1256 (pp. 2-11). Bellingham, WA: International Society for Optical Engineering.

Wickens, C. D., & Todd, S. (1990). Three-dimensional display technology for aerospace and visualization. In Proceedings of the Human Factors Society 34th Annual Meeting, (pp. 1479-1483). Santa Monica, CA: Human Factors Society.

Yeh, Y. Y., & Silverstein, L. D. (1992). Spatial judgments with monoscopic and stereoscopic presentation of perspective displays. Human Factors, 34(5), 583-600.

PROJECTED IMPACT OF A PROTOCOL ADJUSTMENT ON THE INVALID OUTCOME RATE OF THE USAF CYCLE ERGOMETRY ASSESSMENT

Franklin Flatten
Research Associate
Department of Kinesiology and Health Education

Gerald DeWolfe
Research Associate
Department of Kinesiology and Health Education

Stefan Constable, Ph.D.
Exercise Physiologist
Armstrong Laboratories

Lt. Ryan O'Dowd
Air Force Fitness Program

Lt. Dean Richardson
Air Force Fitness Program

The University of Texas at Austin
Department of Kinesiology and Health Education
222 Bellmont Hall
Austin, TX 78712

HSC/YAM
Air Force Fitness Program
8107 13 St.
Brooks AFB, TX 78235-5218

Final Report for:
Graduate Student Research Program
Armstrong Laboratory

Sponsored By:
Air Force Office of Scientific Research
Bolling Air Force Base, DC
and
Armstrong Laboratories

October 1996

PROJECTED IMPACT OF A PROTOCOL ADJUSTMENT ON THE INVALID OUTCOME RATE OF THE USAF CYCLE ERGOMETRY ASSESSMENT

Franklin Flatten, Research Assistant, Department of Kinesiology and Health Education, The University of Texas at Austin

Abstract

Pass, Fail and Invalid outcomes of the US Air Force's Cycle Ergometry Assessment were analyzed from data collected at five AF bases. An Invalid results when the heart rate (HR) response falls outside of the parameters set forth for the assessment (i.e. HR too high or HR below 125 beats per minute (bpm)), the subject requests termination of the assessment, or an error occurs due to either equipment failure or assessment administrator error. Of all tests analyzed 16.4% tests (1548 of 9437) resulted in an Invalid outcome (74.0% Passed, 9.6% Failed). The total Invalid outcomes were then sorted by (seven) categories, and excessive heart rate (HR) i.e., Category 1, accounted for the greatest percentage of Invalids (39.7%). These Invalids are primarily due to the projected workload (WL) being too high. Most subjects who re-test at a lower WL setting receive a score. Therefore, lowering the HR range required for an increase in the WL during the assessment should maintain the HR below the 85% HRmax cutoff and allow for a score to be assessed. Further in depth analysis suggested that a 10-bpm adjustment (decrease) in minutes 3 and 4 of the WL adjustment criteria would potentially reduce the Invalid rate by, at best, only 1.6% of total tests (14.8% total Invalids). This estimate was formulated because most subjects who receive the Category 1 Invalid do not receive any WL progression in minutes 3, 4, or 5. Therefore, no adjustment to the required HR response would affect the WL. The total Invalid rate may be further decreased by other testing protocol adjustments.

INTRODUCTION

The need to accurately assess the fitness level of the Air Force (AF) population has been addressed with a submaximal cycle ergometry (CE) assessment. For a submaximal assessment to predict maximal oxygen consumption ($VO_{2\,max}$) there must be an interval period in which the HR is assessed at steady state. The HR range for the AF CE assessment during this interval is a minimum of 125 bpm, to a maximum of 85% of HR maximum (HRm); i.e. 85% of HRm calculated as [(220 - age) X .85]. If an individuals HR response falls outside of this range, $VO_{2\,max}$ may not be as accurately predicted. The possible outcomes of the AF fitness assessment are Pass, Fail, or Invalid. When an individual's HR response falls outside of the designated range, the assessment is categorized as an Invalid. At present, the CE assessment too often results in an Invalid assessment, specifically Category 1 or high HR, and no "score" is assessed. The subject must then be re-tested on a subsequent day.

Anecdotal evidence from fitness assessment personnel first suggested a majority of Invalid assessments were due to subjects exceeding 85% HRm (Category 1 Invalid). Excessive Invalid assessments and the resulting need for a re-assessment present an unwanted drain on manpower and resources, as well as morale. It was therefore postulated that subjects who received an Invalid Category 1 outcome may have the greatest potential to instead receive a score Pass or Fail) after an adjustment to the workload progression portion of the CE 7-4

protocol is made. Therefore, the purpose of this study was twofold: 1) to determine the rates of Pass, Fail, and especially Invalid assessment outcomes across Categories 1-7, and 2) to analyze the potential impact of a 10 bpm reduction of the heart rate parameters, which determine the workload progression portion of the evaluation, during minutes 3 and 4 on the final assessment outcome (i.e. a Pass, Fail, or Invalid result).

A follow-up study will compare the current assessment protocol to two proposed protocols in an attempt to reduce the overall number of Invalid assessments. The two protocols have been

designated Protocol A and B. Protocol A will alter the computer logic to make it more difficult for a subject to receive a 1.0 kilopond (Kp) or 0.5 Kp workload progression, i.e. lower the minimum HR needed to receive a workload increase (Appendix 1, Part B). Protocol B will lengthen each of the three stages at which workload progression occurs by 1 minute , thereby allowing more time to achieve a steady state HR. Only the potential impact of Protocol A will be discussed further in this report.

METHODS

Available information on fiscal year 1996AF submaximal CE assessments from five bases was collected and analyzed. For this initial evaluation, data is based on the number of total assessments with special interest in service members who received Category 1 Invalid outcomes. These numbers include the same individuals who took repeat assessments. All results are calculated from the combined male and female data unless otherwise noted.

Total assessments (all Pass, Fail, Invalid and re-tests) from Brooks AFB (n=530), Kelly AFB (n=2118), Lackland AFB(n=2191), Patrick AFB (n=2363), and Randolph AFB (n=2235) were sorted and the Pass, Fail, and Invalid rates were determined. Invalid frequencies for the seven Invalid Categories were also calculated. The seven Categories were deliniated by the following:

1) HR exceeds 85% of maximum (HRm; based on 220-age);
2) HR does not reach 125 beats per minute (bpm) in the last minute of the assessment;
3) HR varied more than 3 bpm in the final 2 minutes;
4) Subject could not maintain 50 revolutions per minute (rpm);
5) Rating of Perceived Exertion (RPE) exceeds 15;
6) Subject requested termination of the assessment;
7) Other.

Category 5 (RPE exceeds 15) was deleted in April of 1996.

Combined Category 1 Invalid assessment data from Brooks, Kelly, and Randolph AFB were compiled and further analyzed by HR response and WL progression during minutes 3, 4, and 5 (Tables 2-5). Individual Invalid data from Brooks, Kelly, and Randolph AFB is provided in Appendix 4-7. Due to the large time investment necessary to analyze the data in this manner, this analyses was not done for Lackland or Patrick AFB. However, assessment records for selected individuals from 12 other AF bases who had three Invalid outcomes were also analyzed. These data were separated by Invalid category and only the Category 1 Invalid

assessments were used for separate analyses. These service members WL progressions and HR responses during minutes 3, 4, and 5 were also determined (Table 6).

Re-test data from the five bases were also examined. For this analysis, assessments were separated by subject number so that the total number of subjects could be differentiated from the total number of assessments.

Subject data were downloaded from the FitSoft 2.0 database at four of the bases; while the fifth base, Patrick, uses AF 2000 software (Microfit, Inc.). Protocols and algorithms for Fitsoft 2.0 and AF 2000 are the same regardless of the software. All data were transferred to and sorted on Microsoft Access. Further analysis was performed with Microsoft Excel 5.0.

RESULTS

Invalid assessment outcomes from the five bases accounted for ≈16.4% of all assessments (n=1548 of 9437), while the percentage receiving a Pass was ≈74.0%(n=6985 of 9437), and ≈9.6% Failed (n=904 of 9437; Table 1). The Invalid assessment breakdown as a function of the total number of assessments evaluated is as follows (see Methods for category descriptions): Category 1 accounted for 6.5% of all assessments, Category 2 accounted for 1.3%, Category 3 accounted for 2.5%, Category 4 accounted for 0.08% ,Category 5 accounted for 1.5% (category deleted as of April 1996), Category 6 accounted for 0.3%, and Category 7 accounted for 4.3% of Invalid assessments (Table 1). Again, repeat test outcomes were not discriminated here.

Analysis of only initial assessment outcomes from Brooks, Kelly, and Randolph AFB were completed to determine if the analysis of total combined assessment data was a reasonable approxiamation of what occured on the initial evaluation. Invalid assessment outcomes from the three bases accounted for 16.2% of all assessments (n=660 of 4070), while the percentage receiving a Pass was 75.8%(n=3086 of 4070), and 8.0% Failed (n=324 of 4070; Table 1A). The Invalid assessment breakdown as a function of the total number of assessments evaluated is as follows (see Methods for category descriptions): Category 1 accounted for 6.4% of all assessments, Category 2 accounted for 0.5%, Category 3 accounted for 1.9%, Category 4 accounted for 0.05% ,Category 5 accounted for 2.3% (category deleted as of April 1996), Category 6 accounted for 0.2%, and Category 7 accounted for 4.9% of Invalid assessments (Table 1A).

Combined Category 1 data from Brooks, Kelly, and Randolph AFB were examined by minute of assessment (n=279 for minute 3, n= 264 for minute 4, and n=255 for minute 5) and workload progression (Tables 2, 3 and 4). The number of assessments in each minute declines due to the early termination of some assessments generally because of subjects' HR being greater than 85% of predicted maximum. Category 1 Invalid assessments were reviewed in detail at these bases because the data suggests that changes to the current protocol which impact this category should reduce the greatest number of Invalid assessments (Figure 1). Table

2 shows that 49.8% of these assessments did not receive a workload increase in minute 3. The frequency of not receiving any WL progression increases dramatically in minutes 4 and 5 (81.1% and 72.5%, respectively; Tables 3 and 4). Overall, at roughly 67.4% of the 798 decision points (points during the assessment when a WL progression could occur, i.e., minutes 3, 4, and 5) no WL increase was indicated.

The Invalid assessments from Brooks, Kelly, and Randolph AFB (n=279) were further categorized by the magnitude of the HR response relative to the WL increases during minutes 3, 4, and 5 (Table 5). This breakdown was completed in order to estimate the potential impact of making the WL progression criteria more conservative, i.e. lowering the HR range to make it more difficult to receive a WL progression. Due to the excessive time needed for the analyses, it was not determined if those who received a WL progression in minute 3 also received a WL progression in minute 4 and/or minute 5 or vice versa. Results reported here are based on total assessment data and not on individual responses (the subject is counted as many times as they were re-assessed).

A smaller data base using individuals with three or more Invalid assessments was also used to evaluate the HR response to minutes 3, 4 and 5 of the assessment. The records for 46 subjects were evaluated and it was determined that of 138 assessments, 59 were identified as Category 1 Invalid (Table 6). It was not possible to distinguish between the annual assessment, first re-test, or the second re-test for this data. The data show that 76.3%, 94.8%, and 86.0% of these assessments had no workload progression at minute 3, 4, and 5, respectively. Of the original 59 Invalid assessments, one assessment was terminated before minute 4, and eight were terminated before minute 5. These numbers correspond to a total of 167 decision points. In 143 of these cases (85.6%) no workload progression was received. Any AF member receiving an Invalid must re-take the assessment. Data from the five bases revealed that of subjects who receive a Category 1, 2, 3, 4 or 6 Invalid on their first assessment, 55.8% (n=280) Pass on their first re-test, 22.1% (n=111) Fail, and only 22.1% (n=111) have a second Invalid result (Table 7). Category 5

and 7 Invalid assessments were excluded from the analysis because Category 5 was deleted as an option in April of 1996 and Category 7 is not indicative of subject response, but rather is due to equipment or Fitness Assessment Monitor (FAM) error. Thus, to more accurately evaluate the potential impact of Protocol A, only Invalid categories which are directly caused by or related to the protocol were included in the re-test analysis. Of the 111 individuals with an Invalid outcome on their first re-test, 70 had completed their second re-test with the following results: 44.3% (n=31) Pass, 20.0% (n=14) Fail, and 35.7% (n=25) had a third Invalid assessment. These numbers are only for re-tests after an Invalid and do not include re-tests after a Fail on the first re-test.

DISCUSSION

This study was primarily undertaken because of the perceived high incidence of Invalid fitness assessments due to HR above the accepted range (>85% HRm; Invalid Category 1). Our analysis of available data has shown that Category 1 Invalid assessments account for only 6.5% of all assessments at the five bases studied (n=9437; Table 1). Category 1 Invalid assessment outcomes (n=615) account for 39.7% of all Invalid assessments (Table 1, Figure 1). In other words, even though the percentage of total assessments accounted for by a Category 1 Invalid is lower than expected, the percentage of Invalid Category 1 assessments is still considerable. The impact of a modified protocol on reducing Invalid outcomes will therefore be lower than desired Even so, Protocol A may affect the largest single group of Invalids and therefore have a substantial impact on reducing the total number of Invalid assessments. This protocol change could possibly have some impact on Categories 2-4 and 6 as well. It is speculated that lowering the HR range needed to elicit an increase in WL may increase Category 2 Invalids, but may decrease the number of Category 3, 4 and 6 Invalids.

Protocol A is designed to affect the workload progression by making it more difficult for a subject to receive an increase in workload. For example: a 33 year old subject having a HR of 102 bpm at minute 3 in the current AF protocol would receive a 1 Kp progression, whereas in Protocol A the individual would receive a .5 Kp workload progression (see Appendix 2 for HR criteria), thereby keeping the HR lower. It is estimated that Protocol A could reduce the number of Category 1 Invalids outcomes by only 55.5% at the very best (38.1% of assessments possibly affected in minute 3 + 17.4% of assessments possibly affected in minute 4; Table 5). Therefore, Protocol A could reduce the total number of Category 1 Invalid assessments from 39.7% to 22.7% (From Tables 1 and 5: [(615-341)/(1548-341)](100)=22.7%). A reduction in Category 1 Invalid assessments from 39.7% to 22.7% could reduce the percentage of total Invalid assessments from 16.4% to 12.8%, thus potentially lowering the total number of Invalidassessments by 341 assessments, or 3.6%

It is important to note that multiple assssments (re-tests) by the same subject could not be discerned. Consequently, the predictions presented here are based on the *total* number of assessments. That is, without correction for the possible re-testing of subjects. Therefore, an individual receiving a workload progression in both minute 3 and 4 is evaluated as two assessments. This could easily lead to overestimation of the impact of Protocol A on the Invalid rate(s).

Evaluation of the initial assessment data (n=4070) from Brooks, Kelly, and Randolph AFB, excluding Categories 5 and 7, indicated that 90.2% of these assessments receive a score, 9.8% have an invalid outcome (Table 1A). This percentage was determined by subtracting Category 5 and 7 assessments (n=291) from the total number of Invalid assessments and the number of total assessments (i.e., (660-291) and (4070-291), respectively). 77.9% of individuals with an initial invalid assessment, excluding category 5 and 7, received a score on the first re-test (see Table 7). Therefore, 97.8% of all subjects receive a score within the first two assessments.

Analysis of 279 Category 1 Invalid assessments from Brooks AFB, Kelly AFB and Randolph AFB showed that only 50.2%, 18.9%, and 27.5% (Tables 2, 3, and 4, respectively) of assessments had a workload progression at minutes 3, 4, and 5, respectively. In comparison, the data for subjects with three Invalid assessments (Table 6) show that only 23.7%, 5.2%, and 14.0% of assessments have a workload progression at minutes 3, 4, and/or 5, respectively. This indicates a majority of subjects who receive an Invalid score are riding at or near the initial workload for the entire assessment (see Appendix 2). The initial workload is based on gender, age, weight, and self-reported activity level. While it is possible some individuals could receive a Passing score without an increase in workload, the score would probably indicate that they are in the lowest range of passing scores. As is shown in Table 7, 55.8% of first re-tests result in a Pass while 22.1% Fail. This would indicate that while a preponderance of those who receive an initial Invalid outcome can pass the assessment, it is generally only after the software initiates a lower WL allowing the heart rate to stay lower that they are able to pass (i.e., they are more

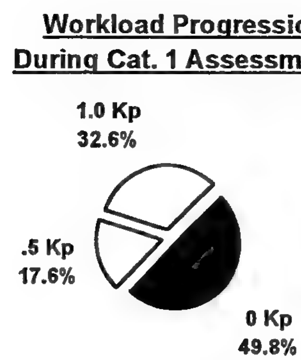
unfit since it takes a lower WL to keep their HR below the upper limit). Since the Fail rate is twice as high in this re-test group compared to the initial assessment outcomes, it appears that the first Invalid outcome is often masking what should be categorized as a Fail. This is an important consideration with regard to further protocol adjustments.

The second largest category of Invalids was Category 7 ("Other"). Data from Brooks, Kelly, Randolph, Lackland, and Patrick AFB's demonstrate that Category 7 Invalids make up 26.5% of all Invalid assessments, and 4.3% of total assessments (Table 1). Category 7 normally indicates FAM error and in very few cases equipment error. Most of the software problems, specific to the assessment, have been identified and corrected with the newest version of FitSoft (FitSoft 2.0), yet computer and equipment failures will continue to happen intermittently. HR monitors may "fail" when the battery runs low, when the monitor is not properly placed during subject preparation, or when the transmitter is too far away from the watch during the assessment. Tester error can include improper HR monitor operation or placement, as well as inaccurate data entry or work-load setting. More thorough training in CE and familiarity and knowledge of the typical responses to exercise may reduce the incidence of Category 7 Invalid outcomes by the FAM. Other factors such as scale calibration (body weight), higher or lower rpm, talking while cycling, self-reported activity level, fan availability, and room temperature all have an undetermined but possible impact on the assessment and may contribute to the high Invalid rate. However, other modest protocol adjustments, i.e. HR variability criteria and computer logic, minimum passing WL criteria and computer logic, etc, may offer the most fruitful and pramatic approach to further reduce the rate of repeated Invalid assessment outcomes.

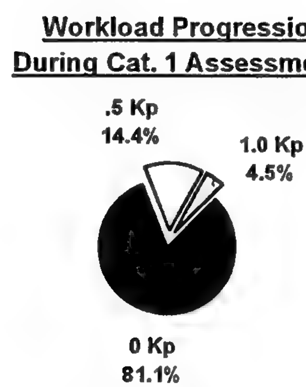
FIGURES

Figure 1: Category 1 Invalid Breakdown By Minute 3, 4, and 5 (Three Bases: Brooks, Kelly, and Randolph AFB [n=279])

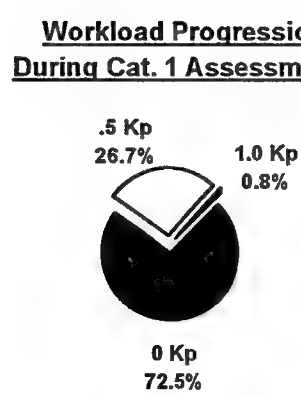
Minute Three:



Minute Four:



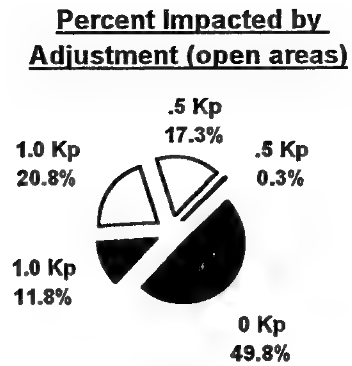
Minute Five:



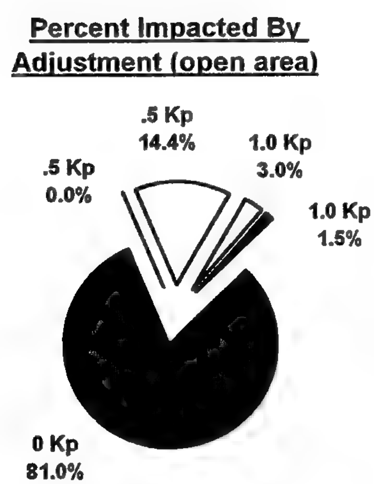
Legend: Open areas: Those that receive a WL increase.

Figure 2: Projected Impact On Category 1 Invalid Tests After A 10 Beat Per Minute Protocol Adjustment (n=279)

Minute Three:



Minute Four:



Minute Five:
No adjustment will be made to minute five.

Best Projected Impact:

- It is estimated that, at best, 55.5% of the Category 1 Invalid assessments can be affected (38.1% + 17.4% of Invalid assessments in minutes 3 and 4 from above).
- Category 1 Invalid assessments could be reduced to 22.7% of total Invalid assessments. This would reduce the number of Invalid assessments to 14.6% of all assessments taken.

Legend: Open areas: Those area impacted by 10 bpm Protocol adjustment.

TABLES

Table 1: Brooks, Kelly, Lackland, Patrick, and Randolph AFB Combined Cycle Ergometry Assessment Breakdown

Brooks, Kelly, Lackland, Patrick, and Randolph AFB Combined Assessment Results			
Test Result	# of assessments	% of Total Tests	
Invalids	1548	16.4	
Fail	904	9.6	
Pass	6985	74.0	
Total:	9437		
Test Result	# of assessments	% of Total Invalids	% of Total Tests
Invalid #1	615	39.7	6.5
Invalid #2	121	7.8	1.3
Invalid #3	233	15.1	2.5
Invalid #4	8	0.5	0.08
Invalid #5	137	8.9	1.5
Invalid #6	24	1.6	0.3
Invalid #7	410	26.5	4.3
Totals:	1548		

Table 1A: Brooks, Kelly, and Randolph AFB Combined Initial Cycle Ergometry Assessment Breakdown

Brooks, Kelly, and Randolph AFB Combined Assessment Results			
Test Result	# of assessments	% of Total Tests	
Invalids	660	16.2	
Fail	324	8.0	
Pass	3086	75.8	
Total:	4070		
Test Result	# of assessments	% of Total Invalids	% of Total Tests
Invalid #1	260	39.4	6.4
Invalid #2	21	3.2	0.5
Invalid #3	79	12.0	1.9
Invalid #4	2	0.3	0.05
Invalid #5	93	14.1	2.3
Invalid #6	7	1.1	0.2
Invalid #7	198	30.0	4.9
Totals:	660		

Table 2: Brooks, Kelly, and Randolph AFB Minute Three Workload Progression of Category 1 Invalid Assessments

	Brooks, Kelly, and Randolph AFB					
	Females		Males		Males and Females	
workload progression	# of assessments	% of total	# of assessments	% of total	# of assessments	% of total
1 Kp	14	20.9	77	36.3	91	32.6
.5 Kp	14	20.9	35	16.5	49	17.6
0 Kp	39	58.2	100	47.2	139	49.8
Total	67		212		279	

Table 3: Brooks, Kelly, and Randolph AFB Minute Four Workload Progression of Category 1 Invalid Assessments

	Brooks, Kelly, and Randolph AFB					
	Females		Males		Males and Females	
workload progression	# of assessments	% of total	# of assessments	% of total	# of assessments	% of total
1 Kp	1	1.6	11	5.5	12	4.5
.5 Kp	4	6.3	34	17	38	14.4
0 Kp	59	92.2	155	77.5	214	81.1
Total	64		200		264	

Table 4 : Brooks, Kelly, and Randolph AFB Minute Five Workload Progression of Category 1 Invalid Assessments

	Brooks, Kelly, and Randolph AFB					
	Females		Males		Males & Females	
workload progression	# of assessments	% of total	# of assessments	% of total	# of assessments	% of total
1 Kp	0	0	2	1	2	0.8
.5 Kp	8	13.1	60	30.9	68	26.7
0 Kp	53	86.9	132	68	185	72.5
Total	61		194		255	

Table 5: Brooks, Kelly, and Randolph AFB Category 1 Invalid Heart Rate Response During CE Assessment

Males and Females Combined							
Work Load Progression (WLP)	beats below initial workload inc.	Minute 3		Minute 4		Minute 5	
		# of assessments	% of total	# of assessments	% of total	# of assessments	% of total
1 Kp	1-5	34	**12.2**	6	**2.3**	1	0.3
	6-10	24	**8.6**	2	**0.7**	1	0.3
	>10	33	11.8	4	1.5	0	
.5 Kp	1-5	18	**6.5**	9	**3.4**	1	0.3
	6-10	30	**10.8**	29	**11.0**	6	2.4
	>10	1		0		61	23.9
	beats above lower limit of WLP						
0 Kp	1-5	7	2.5	25	9.5	14	5.5
	6-10	11	3.9	28	10.6	21	8.2
	11-15	9	3.2	27	10.2	26	10.2
	16-20	18	6.5	13	4.9	24	9.4
	>20	94	33.7	121	45.8	100	39.2
Total		279		264		255	

* Note: Lower limit of workload progression (i.e. highest HR at which an individual can receive a .5 Kp WL progression) determined by age and minute of progression (Appendix 1, Part A).

Table 6: Heart Rate Response During CE Assessment of 46 Individuals with Three Category 1 Invalid Assessments

Males and Females Combined							
Work Load Progression (WLP)	beats below initial workload inc.	Minute 3		Minute 4		Minute 5	
		# of assessments	% of total	# of assessments	% of total	# of assessments	% of total
1 Kp	1-5	4	6.8	0		0	
	6-10	2	3.4	1	1.7	0	
	>10	2	3.4	0		0	
.5 Kp	1-5	3	5.1	0		0	
	6-10	3	5.1	2	3.4	1	2
	>10	0		0		6	12
	beats above lower limit of WLP						
0 Kp	1-5	1	1.7	4	6.9	2	4
	6-10	3	5.1	4	6.9	1	2
	11-15	1	1.7	3	5.2	6	12
	16-20	8	13.6	4	6.9	6	12
	>20	32	54.2	40	69	28	56
Total		59		58		50	

* Note: Lower limit of workload progression (i.e. highest HR at which an individual can receive a .5 Kp WL progression) determined by age and minute of progression (Appendix 1, Part A).

Table 7: First and Second Re-Test Data for Individuals With an Initial Category 1, 2, 3, 4, or 6 Invalid Assessment

Brooks, Kelly, Lackland, Patrick, and Randolph AFB

First Re-Test	result	# of subjects	% of total assessments
	Pass	280	55.8
	Fail	111	22.1
	Invalid	111	22.1
	Total	502	
Second Re-Test	Pass	31	44.3
	Fail	14	20.0
	Invalid	25	35.7
	Total	70	

APPENDIX

Appendix 1

A) Heart rate parameters for workload progression (Protocol B and Original)

	Workload Progression											
	+1 kp			+0.5 kp			0.0 kp			Terminate Assessment		
Minute	3	4	5	3	4	5	3	4	5	3	4	5
Age												
17 - 30	<110	<110	<115	110-119	110-119	115-128	120-173	120-173	129-173	Invalid if >85% of max. heart rate		
31 - 40	<105	<105	<110	105-114	105-114	110-126	115-161	115-161	127-161			
41 - 50	<100	<100	<105	100-109	100-109	105-122	110-152	110-152	123-152			
51 - 60	<100	<100	<105	100-109	100-109	105-120	110-144	110-144	121-144			
61 - 70	<90	<90	<95	90-104	90-104	95-105	105-135	105-135	106-135			

Progression workload cycle changes*.

* Note: Heart rates used to determine workload progression are taken at the end of the minute. For example, minute three of the assessment is performed at the initial workload, with the heart rate at the end of minute three determining the workload progression for minute four using "Minute 3" workload progression column.

B) Heart rate parameters for workload progression (Protocol A)

	Workload Progression											
	+1 kp			+0.5 kp			0.0 kp			Terminate Assessment		
Minute	3	4	5	3	4	5	3	4	5	3	4	5
Age												
17 - 30	<100	<100	<115	100-119	100-119	115-128	120-173	120-173	129-173	Invalid if >85% of max. heart rate		
31 - 40	<95	<95	<110	95-114	95-114	110-126	115-161	115-161	127-161			
41 - 50	<90	<90	<105	90-109	90-109	105-122	110-152	110-152	123-152			
51 - 60	<90	<90	<105	90-109	90-109	105-120	110-144	110-144	121-144			
61 - 70	<80	<80	<95	80-104	80-104	95-105	105-135	105-135	106-135			

Progression workload cycle changes.

AIR FORCE OFFICER QUALIFYING TEST (AFOQT): FORMS Q PRELIMINARY AND OPERATIONAL EQUATING

Theresa M. Glomb
Department of Psychology

University of Illinois at Urbana-Champaign
Department of Psychology
603 East Daniel
Champaign, IL 61820

Final Report for:
Graduate Student Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Armstrong Laboratory

August 1996

AIR FORCE OFFICER QUALIFYING TEST (AFOQT): FORMS Q PRELIMINARY AND OPERATIONAL EQUATING

Theresa M. Glomb
Department of Psychology
University of Illinois at Urbana-Champaign

Abstract

This report is an edited version of the technical report written during the Air Force Office of Scientific Research (AFOSR) Summer Research Program documenting the construction of the AFOQT Forms Q1 and Q2 and the subsequent preliminary and operational equating of these forms to the previous AFOQT Forms P. The full technical report contains three main sections; the first section discusses item selection and the procedures involved in constructing Forms Q, the second section covers the item, subtest and composite level statistics and equating statistics, of the 1993 data collection used for the preliminary equating analyses, and the third section provides this information for the 1995 data used in the operational equating analyses. These equating analyses are integral in linking the new forms of the AFOQT to previous forms to ensure equivalence of measurement and thus, these two sections on the preliminary and operational equating have been retained for discussion in this abbreviated version of the technical report. Results suggest that Forms Q1 and Q2 are sufficiently parallel to one another and equivalent to previous Forms P. Preliminary and operational equating analyses suggest that the cubic smoothing equipercentile equatings are the optimal equatings for each of the five composites on each test form. Using this equipercentile equating with cubic smoothing, preliminary and operational conversion tables were developed and are presented in the full technical report. Readers seeking more extensive coverage of this topic and the discussion of the Forms Q test development effort should consult the full technical report which is currently under review by personnel at Armstrong Laboratory.

AIR FORCE OFFICER QUALIFYING TEST (AFOQT): FORMS Q PRELIMINARY AND OPERATIONAL EQUATING

Theresa M. Glomb

Introduction

The Air Force Officer Qualifying Test (AFOQT) provides aptitude measures for the Air Force's officer selection system. The AFOQT is used to select individuals for Officer Training School, to select Air Force Reserve Officer Training Corps (AFROTC) cadets for the Professional Officers Training School (OTS) and scholarships, and to select students for Undergraduate Pilot Training and Undergraduate Navigator Training. A comprehensive account of the history and development of the AFOQT testing program was authored by Rogers, Roach and Short, 1986.

Periodic updates of the AFOQT to ensure currency, security and predictive validity have historically been the responsibility of the Air Force Human Resources Laboratory (AFHRL), now the Human Resources Directorate of the Air Force's Armstrong Laboratory. Updating the AFOQT begins with the development of parallel test forms that are equivalent to previous AFOQT test forms on item specifications such as statistics and content. In addition to the test development process, updating the AFOQT involves a provisional equating and operational equating study to create conversion tables.

The purpose of the technical report written during the Air Force Office of Scientific Research (AFOSR) Summer Research Program was to describe the construction of the AFOQT Forms Q1 and Q2 and the subsequent preliminary and operational equating of these forms to the previous Forms P. The technical report contains three main sections; the first section discusses item selection and the procedures involved in constructing Forms Q, the second section covers the item, subtest and composite level statistics and equating statistics, of the 1993 data collection used for the preliminary equating analyses, and the third section provides this information for the 1995 data used in the operational equating analyses. These equating analyses are integral in linking the new forms of the AFOQT to previous forms to ensure equivalence of measurement and thus, these two sections on the preliminary and operational equating have been retained for

discussion in this abbreviated version of the technical report. Due to space constraints, this AFOSR report lacks many of the tables of the full technical report which is currently under review by personnel at Armstrong Laboratory. Readers seeking more extensive coverage of this topic and the discussion of the Forms Q test development effort should consult the full technical report.

Preliminary Equating Study

Method

Subjects

Subject samples for the preliminary equating study were selected on availability but also to have a broad range of ability. For this purpose, examinees selected were from samples of the Air Force Academy sophomore and junior class, Air Force ROTC cadets, and airmen from the Basic Military Training School. Hereafter these samples will be referred to as AFA, ROTC, and BMTS respectively. ROTC and BMTS examinees were tested from mid-June to mid-August in 1992. The AFA examinees were tested during the end of the school year in 1993. Demographic frequencies indicate that subjects were predominately male, Caucasian, high school graduates and attained approximately fourteen or fifteen years of education.

Forms Q Test Content

Forms Q were developed to be as similar as possible to previous forms in terms of overall test content, test length, item difficulty, item discrimination, subject matter, and stylistic features. The AFOQT has 380 items comprising 16 subtests which are combined to create five composite scores. The subtest names, the number of items in each subtest and their categorization into the five composites are presented in Table 1. Total testing time, including administrative procedures, is approximately 270 minutes. A more detailed description of the subtest content can be found in the AFOQT Forms P Test Manual (Berger, Gupta, Berger, & Skinner, 1990).

Insert Table 1

Administration

The AFOQT data for the equating study were collected during four and one-half hour testing sessions during which the standardized test procedures were observed as closely as possible. The standardized procedures for administration are provided in the AFOQT Manual For Administration for Forms Q1 and Q2, a document issued by Air Force Personnel Center (AFPC) that explicates standard test conditions, test material preparation, the use of proctors, and the protocol for conducting the testing session. Testing occurred at Lackland Air Force Base for the examinees from the ROTC and BMTS samples and at the Air Force Academy for AFA examinees.

Data Analysis

The data analysis procedures for both the 1993 Preliminary Equating Study and the 1995 Operational Equating Study were nearly identical . Therefore, the data analysis section will be presented only once for the 1993 Preliminary Equating Study, but will serve for the 1995 Operational Equating Study as well. Variations on this data analysis procedure will be noted where appropriate, however, the major difference is that in the 1993 preliminary study, analyses were performed for the subgroups of AFA, ROTC, and BMTS so that future equating efforts will have the opportunity to inform its data collection from previous efforts.

Based on item omitting rates and omit patterns, it was determined that two subtests, Scale Reading and Table Reading, should be analyzed as speeded subtests. For these two subtests, the speeded computational formulas for item statistics were used. The remaining subtests were analyzed as power subtests, even though many have a slight speeded component and would probably be correctly classified as mixed-model subtests.

Classical Item Analysis. Item level data were computed using true score theory (Gulliksen, 1950) item statistics such as item difficulties and item discrimination. Item difficulties (p) are defined as the proportion of examinees who respond correctly to an item. Item difficulties can range from 0.0 to 1.00. Items with difficulties between 0.0 and .30 have a low proportion of respondents answering correctly and are considered hard items. Items with difficulties between .70 and 1.00 have a high proportion of respondents answering correctly and are considered easy items. The reader should note that the term item difficulty is a technical term and seems

contradictory to the lay persons definition of difficulty. An item with a low item difficulty is not an item of low difficulty, but rather a very difficult item.

Biserial correlations (rbis), the correlation between the dichotomously scored item and the continuously distributed subtest score, were computed as measures of item discrimination. Items with discrimination values above .80 are typically viewed as having high discriminatory power; items with discrimination values below .20 are typically viewed as having poor discriminatory power.

Computational formulas for these statistics differ according to whether the subtest is analyzed as a speeded or a power subtest. For a power subtest, item difficulty is calculated using all examinees taking the test, under the assumption that all examinees will have an opportunity to consider every subtest item. For a speeded subtest, difficulty is calculated using only examinees who respond to the item or a subsequent item of the subtest. Examinees who do not attempt items are not considered in these speeded analyses.

Subtest and Composite Analysis. Means, standard deviations, skew, kurtosis, reliability, mean item difficulty and mean biserial correlation values are presented for each subtest. For composite analyses, means, standard deviations, skew and kurtosis values were calculated. Intercorrelation matrices were computed for the subtests and for the composites.

Equating Analysis. Equating enables two forms of a test that are intended to be parallel, which are never precisely equivalent in level and range of difficulty, to be rendered interchangeable by converting the score units of one test to the score units of another. Statistical equating methods establish a relationship between raw scores on two test forms so that the score on one form can be used to express the score on the other form. In the current study, composite scores of Forms Q1 and Q2 were linked to the normative group using linear and equipercentile equating to Forms P scores (see Angoff, 1971 for further explanation of equating).

In linear equating, two raw scores are equated if their z-score values are equivalent, resulting in a smooth straight line. In equipercentile equating, two raw scores are equated if their percentile ranks are equivalent. Because equipercentile equating may result in irregular equating curves, three types of polynomial smoothing (linear, quadratic and cubic) are used, resulting in four possible equatings. The linear and equipercentile equating methods coincide when the score

distributions are the same. In choosing from among the four possible equatings, the z-score linear equating and three polynomial smoothings, the sample descriptive statistics and size are among the characteristics to be considered. When the means, standard deviations, skew, and kurtosis of the two randomly equivalent equating samples are nearly identical on both tests being equated, the z-score linear equating is to be preferred. Linear equating uses two parameters, the mean and standard deviation, per test form. When the z-score linear equating is not appropriate, then one of the three smoothings of equipercentile equatings is chosen. These polynomial smoothings are based upon two parameters for the linear smoothing, three parameters for the quadratic and four parameters for the cubic smoothings. The cubic smoothing of the polynomial equating fits the raw equipercentile data more closely than the quadratic, which fits more closely than the linear. When sample sizes and the range of scores on a test are large, the parameters of the cubic equating are stable and thus, cubic smoothed equipercentile equating should be considered.

Results and Discussion

Item Difficulty Analysis Results

The majority of items in P1 have difficulties ranging from .20 to .80. Electrical Maze is the only subtest that includes items with difficulties below .20. Thirteen of the subtests have at least one item with a difficulty above .80. Approximately half of the items in the Table Reading subtest have item difficulties above .80, suggesting that Table Reading is a relatively easy subtest. Table 2 shows that all sixteen subtests have mean item difficulties between .40 and .60.

Insert Table 2

Form Q1 subtests have similar item difficulty characteristics as subtests in Form P1. Again, item difficulties tend to range from .20 to .80. Two subtests, Electrical Maze and Table Reading have items with item difficulties below .20. Thirteen subtests have at least one item with a difficulty value above .80. Table Reading is a relatively easy subtest; half of the items have

difficulty values above .80. Fifteen subtests have a mean level of item difficulty between .40 and .60.

Item difficulties for test Form Q2 are predominantly in the .20 to .80 range. Three subtests, Verbal Analogies, Mechanical Comprehension and Electrical Maze, include items with item difficulties below .20. Twelve subtests include items with difficulty values greater than .80. As in P1 and Q2, the majority of items from the Table Reading subtest have difficulties above .80. Fifteen subtests have mean levels of item difficulty between .40 and .60.

There are fluctuations in the frequency distributions of the item difficulties on Forms P1, Q1 and Q2. However, these fluctuations could be occurring near the arbitrarily set boundaries for the distribution. There do not appear to be any substantial or systematic differences in the mean item difficulty of a subtest across the three test forms. The maximum difference in subtest mean item difficulty values among any two of the three test forms ranged from .004 to .026. Only four subtests, Arithmetic Reasoning, Reading Comprehension, Scale reading and Hidden Figures, had a largest pairwise difference greater than .020.

Item Discrimination Analyses Results

The items on all three test forms, P1, Q1 and Q2, show acceptable biserial correlations. The frequency distribution of biserial correlations shows that almost all are above .40 and the majority fall in the .60 to .80 range. The subtest mean biserial correlations in Table 2 are generally between .50 and .70 with the minimum mean biserial correlation values of .546, .516. and .536 for Forms P1, Q1, and Q2 respectively. These biserial correlations indicate that the dichotomous item responses correlate well with the subtest score and discriminate well among the examinees.

In comparing the subtest discrimination indices of P1, Q1, and Q2 it is evident that there are fluctuations in the frequency distributions of the biserial correlations. However, these fluctuations could be occurring near the arbitrarily set boundaries for the distribution. There do not appear to be any substantial or systematic differences in the mean biserial correlations of a subtest across the three test forms. The maximum difference in subtest mean biserial correlation values for any two of the three test forms, P1, Q2, and Q2, ranged from .011 to .090. In comparing Forms Q1 and Q2, the difference between subtest mean biserial correlations ranges from .000 to .057.

Subtests Analyses Results

In general, the descriptive statistics of the subtests presented in Table 2 are similar across test forms. Subtests mean scores generally differed by less than one unit. Exceptions to this pattern, or subtest differences greater than one unit were observed between Forms P1 and Q1 on Scale Reading, between Forms P1 and Q2 on Reading Comprehension, Scale Reading and Aviation Information and between Forms Q1 and Q2 on Arithmetic Reasoning. The negligible magnitude of these differences provide support for the parallelism of these measures.

The skew and kurtosis values for the subtests are quite similar across test forms. The majority of the subtests are negatively skewed and none have skew values less than -1.00 or greater than +1.00. Kurtosis values are similar across test forms with a few values around -1.00, a value which indicates a slightly flatter score distribution. Thus, the subtest score distributions are relatively symmetric and tend toward normality.

Kuder-Richardson 20 reliability estimates provide evidence of generally high internal consistency and are approximately equivalent across test forms. The majority of the reliability values are greater than .80, and the lowest estimate is .721. Reliability estimates are not appropriate for subtests scored as speeded tests and thus are not provided for the Scale Reading and Table Reading subtests.

The subtest intercorrelation matrices for each test form show similar correlational patterns. The maximum correlation among subtests is .83, the correlation between Arithmetic Reasoning and Data Interpretation subtests on Form Q2. The minimum correlation is .33 and occurs between the Word Knowledge and Electrical Maze subtests on Form P1 and the Block Counting and Aviation Information subtests on Form Q1. The maximal difference between any of the three subtest correlations in the 120 triads is greater than .10 in only four cases; in these instances the correlation is either .10 or .11. Thus, there is a high degree of similarity among the correlation matrices across the three test forms.

Composite Analyses Results

As would be expected given the similarity in the subtest characteristics, the composite scores are similar across test forms. Composite means for Forms Q1 and Q2 are generally closer than means of P1 with either Form Q. The composite mean scores suggest that Forms Q1 and Q2

are slightly easier than P1, except for the Quantitative composite. Form Q2 has higher mean composite scores than Q1 on the Navigator-Technical, Academic Aptitude, and Quantitative composites while Form Q1 has higher mean composite scores on the Pilot and Verbal composites. However, there should be no significant differences in mean composite scores for Forms Q after the equating.

The skew and kurtosis values for the composites are quite similar across the three test forms. The skew values range from -.38 to -.70; kurtosis values range from -.36 to -.97. These skew and kurtosis values indicate the composite score distributions are relatively symmetric and tend toward normality.

The composite intercorrelation matrices for Forms P1, Q1 and Q2 show that the maximum correlation among composites is .96 and results from the correlation between the Pilot and Navigator-Technical composites on all three forms. The minimum correlation is .75 and occurs between the Verbal and Pilot composites and Verbal and Navigator-Technical composites on Form Q1. The composite intercorrelations are almost identical across test forms; the maximum difference between any of the three composite correlations in a triad is .03. Thus, there is a high degree of similarity among the composite intercorrelation matrices across the three test forms.

Equating Analysis Results

Four possible equatings, the z-score linear, linear smoothed equipercentile, quadratic smoothed equipercentile and cubic smoothed equipercentile, were developed and compared for each composite on Q1 and Q2. The lack of nearly identical moments (skew and kurtosis) for the score distributions ruled out the z-score linear equating method and given that sample sizes were large enough to ensure stability, the cubic smoothing equipercentile equatings were selected for each of the five composites on each test form. Using this equipercentile equating with cubic smoothing, preliminary conversion tables were developed and are presented in the full technical report.

Operational Equating Study

Method

Subjects

Subject samples for the operational equating study were actual examinees taking the AFOQT Forms P1, Q1, and Q2 for purposes of officer selection decisions. Their operational scores were provided by the preliminary conversion tables. These examinees were tested over a period from September of 1994 through June of 1995. On July 1, 1995, Forms Q1 and Q2 were pulled from the field while new equatings were accomplished using applicant scores. Demographic frequencies indicate that subjects were predominately of male gender rather than female gender, Caucasian rather than another ethnicity, with approximately twelve or sixteen years of education and a high school degree as the highest degree earned.

Administration

The AFOQT data for the operational equating study were collected from operational testing sessions at the Military Processing Stations (MEPS) and their outlying sites, Mobile Examining Team Sites (METS). Examiners followed the usual testing procedures for applicants, with the exception that they were to cycle through Forms P1, Q1 and Q2 in that order to all examinees as they came in for testing.

Data Analysis

As mentioned previously, the data analysis procedures for both the preliminary and operational equating studies are similar. The main difference in the two analysis procedures and resultant output is that the preliminary analysis was comprised of total and subsample analyses, whereas the operational analyses involved no subgroup analyses. In addition, the second set of equating analyses, the operational equatings, allowed for comparisons between the preliminary and operational equatings based on the evaluation of critical selection cut areas.

Results and Discussion

Item Difficulty Analysis Results

The majority of items in P1 have difficulties ranging from .20 to .80. Electrical Maze and Table Reading are the only subtests that include items with difficulties below .20. Thirteen of the subtests have at least one item with a difficulty above .80. One-half of the items in the Table Reading subtest have item difficulties above .80, suggesting that Table Reading is a relatively easy subtest. The mean level of item difficulty for the subtests, shown in Table 3, is between .40 and .60 for all sixteen subtests.

Insert Table 3

The item difficulty distributions of subtests Form Q1 are similar to the item difficulty distributions of Form P1. Again, item difficulties tend to range from .20 to .80. Five subtests, Mechanical Comprehension, Electrical Maze, Table Reading, Aviation Information and General Science have items with item difficulties below .20. Twelve subtests have at least one item with a difficulty value above .80. Table Reading is a relatively easy subtest; half of the items have difficulty values above .80. All sixteen subtests have a mean level of item difficulty between .40 and .60.

Item difficulties for test Form Q2 occur predominantly in the .20 to .80 range. Six subtests, Verbal Analogies, Mechanical Comprehension, Electrical Maze, Table Reading, Aviation Information and General Science, include items with item difficulties below .20. Eleven subtests include items with difficulty values greater than .80. As in P1 and Q2, the majority of items form the Table Reading subtests have difficulties above .80. Fifteen subtests had mean level of item difficulty between .40 and .60.

The subtest difficulties of P1, Q1, and Q2 show fluctuations in the frequency distributions of the item difficulties, however, these fluctuations could be occurring near the arbitrarily set boundaries for the distribution. There do not appear to be any substantial or systematic differences in the mean item difficulty of a subtest across the three test forms,. The maximum difference in

subtest mean item difficulty among any two of the three test forms ranged from .002 to .034. Only four subtests, Arithmetic Reasoning, Reading Comprehension, Scale Reading and Hidden Figures, had a largest pairwise difference above .020.

Item Discrimination Analyses Results

The items on all three test forms, P1, Q1 and Q2, show acceptable biserial correlations. The frequency distribution of biserial correlations shows that the majority of the item biserial correlations fall in the .40 to .80 range. The subtest mean biserial correlations in Table 3 are generally between .50 and .70 with the minimum mean biserial correlation values of .511, .490. and .523 for Forms P1, Q1, and Q2 respectively. These biserial correlations indicate that the dichotomous item responses correlate well with the subtest score and discriminate well among the examinees.

Comparisons of the subtest discrimination indices of P1, Q1, and Q2 show that there are fluctuations in the frequency distributions of the biserial correlations. However, these fluctuations could be occurring near the arbitrarily set boundaries for the distribution. There do not appear to be any substantial or systematic differences in the mean biserial correlations of a subtest across the three test forms. The maximum difference in subtest mean biserial correlation values for any two of the three test forms, P1, Q2, and Q2, ranged from .016 to .068. In comparing Forms Q1 and Q2, the difference between subtest mean biserial correlations ranges from .000 to .046.

Subtests Analyses Results

In general, the descriptive statistics of the subtests presented in Table 3 are similar across test forms. Subtests mean scores generally differed by less than one unit. Exceptions to this pattern, or subtest differences greater than one unit were observed between Forms P1 and Q1 on Scale Reading, between Forms P1 and Q2 on Reading Comprehension, Scale Reading and General Science and between Forms Q1 and Q2 on Arithmetic Reasoning and Scale Reading. The negligible magnitude of these differences provide support for the parallelism of these measures.

The skew and kurtosis values for the subtests are quite similar across test forms. The majority of the subtests are negatively skewed and none have skew values less than -1.00 or greater than +1.00. Kurtosis values are similar across test forms with a few values around -1.00, a

value which indicates a slightly flatter score distribution. Thus, the subtest score distributions are relatively symmetric and tend toward normality.

Kuder-Richardson 20 reliability estimates provide evidence of generally high internal consistency and are quite similar across test forms. The majority of the reliability values are greater than .80, and the lowest estimate is .685. In general, these reliability values are lower than those obtained in the preliminary equating study. Reliability estimates are not appropriate for subtests scored as speeded tests and thus are not provided for the Scale Reading and Table Reading subtests.

The subtest intercorrelation matrices of Forms P1, Q1 and Q2 show the maximum correlation among subtests is .76, the correlation between Arithmetic Reasoning and Data Interpretation subtests on Form Q2. The minimum correlation is .20 and occurs between the Word Knowledge and Electrical Maze subtests on Form P1. The subtest intercorrelations show similar patterns across the three forms. The maximal difference between any of the three subtest correlations in the 120 triads is greater than .10 in only two cases; in these instances the correlations are .10 and .11. Thus, there is a high degree of similarity among the correlation matrices across the three test forms.

<u>Composite Analyses Results</u>

As would be expected given the similarity in the subtest characteristics, the composite scores are similar across test forms. Composite means for Forms Q1 and Q2 are generally closer than means of P1 with either Form Q1 or Q2. The composite mean scores suggest that Forms Q1 and Q2 are slightly easier than P1, except for the Quantitative composite. Form Q2 has higher mean composite scores than Q1 on all composites, however, there should be no significant differences in mean composite scores for Forms Q after the equating.

The skew and kurtosis values for the composites are quite similar across the three test forms. The skew values range from -.14 to -.28; kurtosis values range from -.10 to -.80. These skew and kurtosis values indicate the composite score distributions are relatively symmetric and tend toward normality.

The composite intercorrelation matrices for Forms P1, Q1 and Q2 show that the maximum correlation among composites is .93 and results from the correlation between the Pilot

and Navigator-Technical composites on all three Forms. The minimum correlation is .60 and occurs between the Pilot and Verbal composites on Form P1. The composite intercorrelations are nearly identical across test forms; the maximum difference between any of the three composite correlations in a triads is .02. Thus, there is a high degree of similarity among the composite intercorrelation matrices across the three test forms.

Equating Analysis Results

Four possible equatings, the z-score linear, linear smoothed equipercentile, quadratic smoothed equipercentile and cubic smoothed equipercentile, were developed and compared for each composite on Q1 and Q2. As was the case in the preliminary equating study, the evaluations of the equatings ruled out the z-score linear equating and given that sample sizes were large enough to ensure stability, the cubic smoothing equipercentile equatings were selected for each of the five composites on each test form. Using this equipercentile equating with cubic smoothing, operational conversion tables were developed and are presented in the full technical report.

Implementation Effects of Instituting the Operational Conversion Tables

The preliminary conversion tables were used during the selection and classification of officer commissioning applicants during the data collection for the operational equating study. The data from the operational equating study were used to develop the operational equating tables, which were not identical to the preliminary conversion tables. Minor discrepancies in the conversion tables were expected due to the differences in the samples used for the preliminary and operational equatings. The sample of officer commissioning applicants used in the operational equating was larger and more motivated than that used in the preliminary equating study, and thus equatings developed on this sample were preferable. However, it was important to determine if the introduction of the operational tables would cause significant changes in qualification rates for officer positions. Qualification is determined by minimum cut-off values on some or all AFOQT composites for occupational categories such as pilot, navigator, missile, technical and non-line officers depending on the commissioning source of AFROTC, OTS, or the Airmen Enlisted Commissioning Program (AECP).

To examine the effects of the operational conversion tables, the various minimum cut-off values for officer categories and commissioning sources were identified and the raw score conversions to percentiles for both the preliminary and operational conversion tables were listed for a range of percentiles about those minimum. The two conversion tables were very close except for the Navigator-Technical composite on Form Q2 at the tenth percentile. ROTC pilot qualification requires a minimum percentile of 50 on the Pilot composite and a 10 on the Navigator-Technical composite for applicants without a pilot's license and requires a minimum percentile of 25 on the Pilot composite with a 10 on the Navigator-Technical composite for applicants with a pilot's license. A distribution of applicants in the operational equating sample with Pilot Composite scores of 50 through 59 (n=367) showed none with a Navigator-Technical score anywhere as low as the tenth percentile. A distribution of applicants in the operational equating sample with Pilot composite scores of 25 through 34 (n=352) found only three cases with a Navigator-Technical percentile less than 10 and only 8 cases with a Navigator-Technical percentile less than 15. Therefore, the tenth percentile minimum is basically an irrelevant minimum, so there will be no noticeable operational effect in switching from the preliminary conversion tables to the operational conversion tables.

Conclusions and Recommendations

The AFOQT Forms Q1 and Q2 operational conversion tables based on the operational equating study should be implemented for use in making officer selection decisions. The operational conversion tables are more acceptable than the preliminary conversion tables because they were based on the responses of the larger, more appropriate sample used in the operational equating study. In the operational equating study the subjects were actual applicants for officer commissioning who were motivated to do well, thus the operational conversions tables developed on this sample are preferable.

References

Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), Educational Measurement (2nd ed.). Washington, DC: American Council on Education.

Berger, F. R., Gupta, W. B., Berger, R. M., & Skinner, J. (1990). Air Force Officer Qualifying Test (AFOQT) Form P: Test Manual (AFHRL-TR-89-56, AD-A221 004). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Gulliksen, H. (1950). Theory of mental tests. New York: John Wiley & Sons, Inc.

Rogers, D. L., Roach, B. W., & Short, L. O. (1986). Air Force Officer Qualifying Test Form O: Development and Standardization (AFHRL-TR-86-24, AD-A172 037). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Table 1

Description of AFOQT Forms Q Subtests and Composition of Aptitude Composites

Subtest		Number of items	Testing time (minutes)	Composites				
				Pilot	Nav-Tech	Acad. Apt.	Verbal	Quant.
Verbal Analogies	(VA)	25	8	X		X	X	
Arithmetic Reasoning	(AR)	25	29		X	X		X
Reading Comprehension	(RC)	25	18			X	X	
Data Interpretation	(DI)	25	24		X	X		X
Word Knowledge	(WK)	25	5			X	X	
Math Knowledge	(MK)	25	22		X	X		X
Mechanical Comprehension	(MC)	20	22	X	X			
Electrical Maze	(EM)	20	10	X	X			
Scale Reading	(SR)	40	15	X	X			
Instrument Comprehension	(IC)	20	6	X				
Block Counting	(BC)	20	3	X	X			
Table Reading	(TR)	40	7	X	X			
Aviation Information	(AI)	20	8	X				
Rotated Blocks	(RB)	15	13		X			
General Science	(GS)	20	10		X			
Hidden Figures	(HF)	15	8		X			
Total		380	208[a]					

Note. [a] This testing time is for minutes actually spent on the test items. Total test time including administrative activities is 270 minutes.

Table 2

Statistics of Subtests for Preliminary Equating Study

Test	Mean item difficulty			Mean biserial correlation			Mean subtest score			Standard deviation			Skew			Kurtosis			Reliability		
	P1	Q1	Q2	P1	Q1	Q2	P1	Q1	Q2	P1	Q1	Q2	P1	Q1	Q2	P1	Q1	Q2	P1	Q1	Q2
VA	.587	.599	.593	.644	.634	.648	16.83	17.39	17.72	4.90	4.66	4.72	-.63	-.72	-.88	-.38	-.10	.36	.842	.828	.833
AR	.563	.551	.574	.733	.710	.718	15.94	15.38	16.63	6.31	6.24	6.28	-.30	-.25	-.44	-1.07	-1.04	-.97	.907	.899	.908
RC	.546	.560	.571	.668	.648	.705	15.11	15.87	16.51	5.89	5.72	6.13	-.40	-.48	-.59	-.77	-.68	-.70	.882	.870	.896
DI	.570	.561	.566	.603	.659	.693	16.37	15.98	16.28	5.19	5.70	6.12	-.43	-.52	-.56	-.73	-.70	-.81	.841	.873	.892
WK	.559	.564	.550	.707	.709	.655	15.71	15.98	15.26	6.10	6.00	5.54	-.46	-.55	-.44	-.87	-.67	-.68	.897	.894	.869
MK	.588	.592	.596	.822	.811	.848	17.45	17.62	17.89	6.91	6.62	6.94	-.64	-.73	-.78	-.97	-.80	-.73	.931	.926	.936
MC	.491	.501	.494	.565	.598	.592	9.55	10.06	9.76	4.30	4.52	4.41	.17	-.02	.07	-.80	-1.02	-.88	.786	.812	.807
EM	.451	.454	.466	.546	.516	.536	7.81	7.99	8.56	3.79	3.64	3.75	.53	.30	.23	.09	-.27	-.37	.745	.721	.743
SR	.563	.586	.584	.581	.636	.650	25.30	27.43	27.23	7.86	8.10	8.30	-.32	-.71	-.62	-.73	-.26	-.51			
IC	.545	.556	.544	.758	.757	.778	12.12	12.56	12.08	5.61	5.43	5.76	-.36	-.36	-.33	-1.10	-1.10	-1.19	.900	.896	.908
BC	.585	.583	.567	.703	.632	.624	13.53	13.51	12.86	4.46	4.05	4.20	-.62	-.65	-.37	-.19	-.10	-.53	.855	.812	.814
TR	.598	.601	.606	.727	.716	.716	27.65	28.00	28.40	8.07	7.92	8.01	-.47	-.68	-.58	-.29	.13	-.24			
AI	.497	.504	.509	.634	.648	.674	9.32	10.18	10.42	4.49	4.87	5.03	.12	.17	.15	-.93	-.96	-1.10	.829	.848	.862
RB	.537	.548	.538	.602	.667	.653	8.59	9.09	8.73	3.38	3.45	3.40	-.07	-.42	-.30	-.84	-.60	-.62	.782	.786	.777
GS	.512	.516	.516	.608	.596	.601	9.97	10.64	10.73	4.17	4.26	4.50	-.07	-.19	-.16	-.84	-.70	-.76	.799	.795	.814
HF	.595	.569	.586	.723	.705	.734	10.44	9.65	10.21	3.40	3.48	3.60	-.43	-.17	-.45	-.66	-.88	-.69	.814	.813	.831

Note. VA = Verbal Analogies; AR = Arithmetic Reasoning; RC = Reading Comprehension; DI = Data Interpretation; WK = Word Knowledge; MK = Math Knowledge; MC = Mechanical Comprehension; EM = Electrical Maze; SR = Scale Reading; IC = Instrument Comprehension; BC = Block Counting; TR = Table Reading; AI = Aviation Information; RB = Rotated Blocks; GS = General Science; HF = Hidden Figures.

Table 3

Statistics of Subtests for Operational Equating Study

	Mean item difficulty			Mean biserial correlation			Mean subtest score			Standard deviation			Skew			Kurtosis			Reliability		
Test	P1	Q1	Q2	P1	Q1	Q2	P1	Q1	Q2	P1	Q1	Q2	P1	Q1	Q2	P1	Q1	Q2	P1	Q1	Q2
VA	.567	.582	.583	.520	.561	.544	15.76	16.49	16.71	4.34	4.28	4.20	-.30	-.32	-.51	-.37	-.36	-.03	.779	.781	.774
AR	.545	.535	.558	.664	.647	.646	14.87	14.43	15.64	5.73	5.73	5.47	-.08	.03	-.12	-.88	-.90	-.86	.877	.871	.865
RC	.541	.552	.569	.585	.574	.568	14.79	15.45	16.35	5.10	5.11	4.91	-.16	-.19	-.30	-.66	-.73	-.63	.833	.827	.820
DI	.562	.555	.561	.515	.566	.563	15.84	15.57	15.91	4.48	4.94	4.94	-.25	-.27	-.29	-.41	-.62	-.66	.770	.818	.815
WK	.548	.553	.543	.649	.641	.595	15.11	15.39	14.84	5.63	5.46	5.08	-.15	-.21	-.11	-.89	-.78	-.74	.870	.863	.835
MK	.572	.577	.587	.689	.674	.693	16.53	16.68	17.28	5.97	5.61	5.66	-.36	-.31	-.48	-.90	-.85	-.72	.888	.877	.883
MC	.475	.476	.477	.537	.593	.581	8.82	8.87	8.98	4.07	4.48	4.32	.33	.25	.31	-.60	-.91	-.69	.755	.810	.797
EM	.438	.439	.451	.518	.490	.523	7.33	7.33	7.92	3.55	3.41	3.60	.44	.35	.41	.03	-.15	.05	.714	.685	.723
SR	.540	.559	.568	.511	.558	.546	23.29	24.99	25.64	6.92	7.45	7.19	-.15	-.32	-.31	-.23	-.33	-.37			
IC	.515	.517	.511	.684	.705	.692	10.71	10.80	10.52	5.27	5.32	5.30	.08	.09	.10	-1.13	-1.18	-1.10	.872	.880	.875
BC	.563	.562	.550	.672	.624	.604	12.70	12.62	12.14	4.44	4.09	4.17	-.48	-.40	-.32	-.38	-.20	-.40	.841	.807	.801
TR	.594	.594	.603	.681	.636	.640	27.21	27.26	27.85	7.19	7.01	6.93	-.56	-.45	-.51	.63	.47	.53			
AI	.458	.461	.467	.590	.641	.640	7.57	8.17	8.47	4.01	4.66	4.59	.90	.85	.86	.21	-.01	.01	.786	.836	.831
RB	.512	.527	.518	.648	.644	.630	7.77	8.39	8.01	3.32	3.40	3.29	-.03	-.21	-.11	-.67	-.73	-.69	.770	.774	.756
GS	.487	.496	.499	.554	.548	.538	8.87	9.71	9.95	3.86	3.87	3.98	.24	.19	.11	-.50	-.45	-.58	.753	.750	.751
HF	.577	.543	.564	.674	.643	.643	9.79	8.82	9.51	3.20	3.26	3.28	-.29	.01	-.15	-.43	-.57	-.69	.776	.768	.772

Note. VA = Verbal Analogies; AR = Arithmetic Reasoning; RC = Reading Comprehension; DI = Data Interpretation; WK = Word Knowledge; MK = Math Knowledge; MC = Mechanical Comprehension; EM = Electrical Maze; SR = Scale Reading; IC = Instrument Comprehension; BC = Block Counting; TR = Table Reading; AI = Aviation Information; RB = Rotated Blocks; GS = General Science; HF = Hidden Figures.

USE OF THE UNIVERSAL GENECOMB™ ASSAY TO DETECT *ESCHERICHIA COLI* 0157:H7

Leigh K. Hawkins
Graduate Student
Department of Horticulture

Auburn University
101 Funchess Hall
Auburn, AL 36849

Final Report for :
Graduate Student Research Program

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Armstrong Laboratories

August 1996

USE OF THE UNIVERSAL GENECOMB ASSAY TO DETECT *ESCHERICHIA COLI* 0157:H7

Leigh K. Hawkins
Graduate Student
Department of Horticulture
Auburn University

ABSTRACT

The Universal *GeneComb*™ test kit from BioRad is based on DNA hybridization and is used for the rapid detection of PCR-amplified biotin-labeled DNA. Using very specific probe sequences, this kit may be used to detect the 60 megadalton (MDa) plasmid and the Shiga-like toxins (SLTs) of *Escherichia coli* O157:H7. This method proved to be simple and effective. It allowed for a rapid analysis of the PCR-amplified DNA.

USE OF THE UNIVERSAL GENECOMB ASSAY TO DETECT
ESCHERICHIA COLI 0157:H7

Leigh K. Hawkins

INTRODUCTION

Enterohemorrhagic *Escherichia coli* (EHEC) O157:H7 causes diarrhea, hemorrhagic colitis, hemolytic-uremic syndrome (HUS), and thromobocytopenia purpura. *E. coli* O157: H7 isolates have been implicated in major foodborne outbreaks worldwide. Most recently in Japan, this pathogen has been cited as the cause of the worst outbreak of food poisoning in a decade (Anonymous, 1996). Cattle are considered the major reservoir of O157:H7 in the U.S. and most infections in the United States result from eating undercooked hamburger meat. EHEC pathogens have also been transmitted in dry fermented sausage, milk, apple cider, mayonnaise, various salads, water, and by direct person-to-person and cattle-to-person contacts (Abbot et. al., 1994).

O157:H7 isolates produce cytotoxins known as Shiga-like toxins (SLT). The existence of two major types of SLT has been demonstrated on the basis of antigenic and nucleotide sequence variations (Bettelheim et. al., 1993). SLT I and SLT II are immunologically distinct and demonstrate a 58% nucleotide and 56% amino acid sequence homology (Acheson et. al., 1994; Brian et. al., 1992; Jackson et. al., 1987). SLT I resembles the Shiga toxin of *Shigella dysenteriae* type 1 in amino acid sequence, structure, and activity. *E. coli* O157:H7 isolates also possess a 60 megadalton (MDa) plasmid, which is associated with the pathogenicity of EHEC (Cebula et. al., 1995; Levine et. al., 1987).

The objective of my research this summer was to evaluate the effectiveness of the Universal GeneComb ™ Test Kit (BioRad, Hercules, CA) for rapidly detecting the presence of the PCR-amplified regions of the 60 MDa plasmid and SLTs I and II of Escherichia coli O157:H7. The test kit is used for the rapid detection of PCR-amplified DNA. This method can be used instead of agarose gel electrophoresis to detect PCR amplicons. Using very specific DNA probe sequences, the GeneComb can distinguish between the two SLT toxins. Agarose gel electrophoresis does not provide enough resolution to distinguish between the SLT I and SLT II amplicons because the fragment sizes do not differ much, 227 and 224 bp.

DNA hybridization is a highly sensitive and specific assay that has been used as a diagnostic technique for identification of various microorganisms (Reinhartz et. al., 1993). This technique is employed by the Universal Gene Comb from BioRad. The GeneComb assay is based on the chromatographic migration of PCR-amplified, biotin-labeled DNA on a nitrocellulose strip passing through an immobilized probe area. This method uses capillary action to bring the DNA into contact with the immobilized probe. DNA with a sequence homologous to the probe hybridizes to it and is detained in the probe area. Unhybridized DNA continues to migrate away from the hybrid area. The biotinylated hybrid is visualized by a color reaction using streptavidin-alkaline phosphatase (SA-AP) conjugate and a chromagenic substrate (Reinhartz et. al., 1993 and manufacturer's directions).

METHODOLOGY

Several strains of *Shigella and Escherichia* including several serotypes were used in this study. The strains were clinical isolates from the Centers of Disease Control and Prevention (Atlanta, GA), the Alabama Department of Health (Montgomery, AL), the Texas Department of Health (Austin, TX), and Brooks Air Force Base Armstrong Laboratories (San Antonio, TX) and bovine isolates from the Auburn University School of Veterinary Medicine (Auburn, AL). All strains were stored at -70°C.

Several tests were employed to identify and characterize the strains. The Vitek GNI (bioMérieux Vitek, Inc., Hazelwood, MO) test performs 30 different biochemical reactions for use in identifying gram negative bacteria. The reaction which is of most interest is the sorbitol test. The lack of sorbitol fermentation is a characteristic of O157:H7 and is used to isolate the bacteria from clinical and food specimens (Abbott et. al., 1994, Cebula et. al, 1995; Gunzer et. al., 1992). The sorbitol negative response was confirmed by growth on the differential media MacConkey agar and MacConkey agar with sorbitol (Remel, Lenexa, KS). A color change indicates a positive response due to acid production. This is not a definitive test because some O157:H7 strains do ferment sorbitol and other serovars of *E. coli* have been found to be sorbitol negative (Acheson, 1996; Cebula, 1995). The Oxoid *E. coli* O157 kit (Unipath Limited, Hampshire, England), a latex agglutination test, was used to detect the presence or absence of the O157 antigen. This test uses latex particles coated with rabbit antibodies to agglutinate those bacterial

strains which possess the O157 antigen. The latex agglutination test also is not a definitive test. Other *Escherichia* serovars and a *Citrobacter freundii* strain have been found to cross-react with the O157 antiserum (Rice et. al., 1992; Bettelheim et. al., 1993). The Premier EHEC enzyme immunoassay (Meridian Diagnostics, Inc., Cincinnati, OH) was used for detection of the Shiga-like toxins I and II. A list of the strains used and the above test results are given in table 1.

Table 1: Bacterial Strains and Biochemical Test Results

ID #	Strain	Sorbitol	O157 Antigen	EHEC	NOTES
A5	*Escherichia coli* O157:H7	-	+	+	CDC A7793
A11	*E. coli* O157: H7	-	+	+	CDC C8958
A13	C600:933J	-	+	+	
A14	C600:933W	-	+	+	
A21	*E. coli* O157: H7	-	+	+	
A40	V517	+	-	-	
A45	*E. coli* O157: H7	-	+	+	
A57	*E. coli* O157: H7	-	+	-	ATCC 43888
A58	*E. coli* O157: H7	-	+	+	ATCC 43890
A59	*E. coli* O157: H7 pZC373	-	+	+	
A63	*E. coli* O157: H7	+	+	-	
A98	*E. coli* O157: H7	-	+	+	BE3-1676
A112	*E. coli* O157: H7	+	+	+	BE4-1269
A123	DH5α	+	-	-	
A124	*E. coli* O157:H7	-	+	+	BT96 2935
A135	*E. coli* O44	+	-	-	CDC4756-59
A136	*Shigella dysenteriae* type 3	-	-	-	
A139	*Shigella sonnei*	-	-	-	

O157:H7 strain A45 served as the positive control for the plasmid and SLT I and SLT II toxin genes. A63 and A112 are both atypical O157:H7 strains, both are sorbitol positive and A63 is EHEC negative. A5, A11, A21, A112, A124 are all typical O157 strains. A57 and A58 both contain the plasmid, but A57 produces neither toxin nor does it possess the genes to do so and A58 produces only SLT I (Gherna et. al., 1992). A13 and A14 are non-O157:H7 *E. coli* strains containing the toxin-converting phages 933J (SLT I) and 933W (SLT II) respectively (Jackson et. al., 1987; Strockbine et. al., 1986). Neither A13 nor A14 possess the 60 MDa plasmid. A59 is an O157:H7 strain which has been cured of its plasmid (unpublished data) and served as the plasmid negative control. A123 (DH5α) served as the plasmid and SLT I and II negative control.

The primers MK1, MK2, BFS1R, and BFS1F (table 2) were used to amplify the plasmid and the SLT I and II toxin sequences (Fratamico et. al., 1995). Only one primer pair (MK1 and MK2) is needed to amplify the DNA sequences for both SLT I and SLT II (Karch and Meyer, 1989). Multiplex PCR allows for the simultaneous amplification of the plasmid and SLT toxin genes. The conditions for the multiplex PCR were worked out by another student, Catherine A. Ramaika. The manufacturer (Biorad, Hercules, CA) states that the PCR-amplified product must be biotinylated in order for the GeneComb assay to work. This is easily accomplished by having at least one member of each primer pair biotinylated by the addition of a 5'-biotin-labeled T residue. Biotinylation had no effect on the PCR results when analyzed by agarose gel electrophoresis. Without biotinylation of at least one primer, the product cannot be visualized by the GeneComb assay.

Table 2: Primers and Sequences

PRIMER PAIRS	SEQUENCE (5'→3')	TARGET
MK1 MK2	TTT ACG ATA GAC TTC TCG AC 5'-(BioT)* -CAC ATA TAA ATT ATT TCG CTC	60 MDa plasmid
BFS1F BFS1R	5'-(BioT)*-CTT CAC GTC ACC ATA CAT AT 5'-(BioT)*-ACG ATG TGG TTT ATT CTG GA	SLT I and SLT II

*BioT=biotinylated

A single bacterial colony from an overnight culture on brain heart infusion (BHI) agar was suspended into 200µl lysis solution (0.5% Triton X-100, 20mM Tris (pH 8.0), and 2mM EDTA) and heated at 100°C for 10 minutes to lyse the cells. The 100µl PCR reaction volume consisted of 5 µl crude cell lysate, 2.0mM $MgCl_2$, 10mM Tris (pH 8.0), 50mM KCl, 200µM each of the four deoxyribonucleic triphosphates, 2.5 U *Taq* DNA polymerase (Perkin Elmer, Branchburg, NJ) and 50 pmol of each primer (The Midland Certified Reagent Company, Midland, TX) . The reaction mixture was brought up to volume with water. The amplification was carried out in an automated thermal cycler (Perkin-Elmer GeneAmp PCR System 9600) using the following conditions: an initial denaturation at 94°C for 5 minutes

followed by 35 cycles of denaturation (1 min at 94°C), annealing (3 min at 48°C), and extension (4 minutes at 72°C), and a final extension at 72°C for 5 minutes. The PCR product was then stored at 4°C.

The GeneComb was employed to detect the specific amplified DNA sequences. Also, the PCR product (166 bp for the plasmid and 227/224 bp for SLT I/II respectively) was visualized by ethidium bromide staining of a 1.6% agarose gel. In order to use the comb, a probe at least 20 bp in length and designed to obtain maximum specificity must be created. Our probes were MFSPP for the plasmid (Fratamico et.al., 1995), MKP1 for SLT I, and MKP2 for SLT II (Karch and Meyer, 1989). (Table 3). Each probe is diluted 1:10 (v/v) in freshly prepared binding buffer, prepared from reagents in the test kit, for a final concentration of at least 5 pmol/µl.

Table 3: Probes and Sequences

PROBE	SEQUENCE	TARGET
MFSPP	CCG TAT CTT ATA ATA AGA CG	60 MDa plasmid
MKP1	GAT AGT GGC TCA GGG GAT AA	SLT I
MKP2	AAC CAC ACC CAC GGC AGT TA	SLT II

The GeneComb has eight nitrocellulose teeth including a control tooth to which two probes have been pre-applied by the manufacturer. The lower probe on the control tooth is a random oligonucleotide and the upper probe is an human leukocyte antigen (HLA) sequence. The control sample is an amplified HLA sequence. The control is used to determine the validity of that particular assay . Absence of a spot in the upper probe area and /or the presence of a spot in the lower probe area invalidates the test.

On each tooth, one or two probes may be used for evaluation of the PCR product. For a single determination, 0.5µl of the diluted probe is loaded into the center of the tooth. For a dual determination, 0.5 µl of each diluted probe is loaded diagonally across from each other onto the tooth. We have also found that a mixture of two probes can be used to detect the presence of one or both amplified regions of DNA. In short, equal volumes of MKP1 and MKP2 were mixed together and 0.5µl of the mixture was spotted onto a tooth. This spot is able to pick up either SLT I, SLT II, or both toxins, thus increasing the

capacity of the comb in analyzing the amplicons. The probe(s) is then fixed onto the nitrocellulose teeth by a 3 minute exposure to UV light. With each assay, a negative sample was run to determine the possible background noise as recommended by the manufacturer.

The comb is moved sequentially through four rows of microwells containing different reagents. The amplified DNA product is denatured using the reagents in the kit. The reaction is neutralized by adding another reagent, HybriRun. Then 50µl of each denatured, neutralized sample is put into a microwell in the first row. The Genecomb is incubated in this row for 15 minutes at 37°C so that the DNA can migrate up the tooth and hybridize to the probe area. The comb is then transferred to the second row containing the streptavidin-alkaline phosphatase conjugate and allowed to incubate for 5 minutes at room temperature. The color reaction occurs when the comb is transferred to the third row containing the chromagenic substrate and allowed to incubate for seven minutes at room temperature. The fourth row of microwells contains the stop reaction mix and the comb incubates here for 3 minutes at room temperature. In total, the assay itself takes 30 minutes. A blue-purple spot in the area of the probe is a positive reaction indicating sequence homology between the probe and the amplified DNA product. If there is no spot, then the reaction is considered negative.

RESULTS and DISCUSION

The biochemical and serological asays of the bacterial strain gave results which were consistent with what was expected with exception of strain A63. A63 was thought to an O157:H7 strain, but it gave atypical responses to most of the biochemical and serological tests (table 1). When subjected to the PCR and subsequent analysis by electrophoresis and the GeneComb, A63 tested negative for the presence of the plasmid and both SLT toxin strains. A63 was further characterized by Catherine A. Ramaika by examination by miniscreen and analyzis by field inversion gel electrophoresis (FIGE). When compared to other reference strains, A63 showed many bands which were not related. Further testing will be conducted on this strain.

There was another anomaly when the strains were analyzed by electrophoresis and the GeneComb. Strain A5 is missing the 60 MDa plasmid. A5 was examined by FIGE. No large plasmid

(ca. 80-100kb) was observed. It is not known if this strain ever had a plasmid or lost it upon cultivation in the lab. More research will be dedicated to this strain in future studies.

Figure 1 shows a completed comb and the corresponding agarose gel is pictured in figure 2. The upper spot on each tooth of the comb is MFSPP and the lower spot is a 1:1 (v/v) mixture of MKP1 and MKP2. The same information is given by the gel. On the agarose gel, the 166 bp band is the plasmid fragment and the 224/227 bp band is the fragment of the SLT toxins.

In some cases, we have found that there has been nonspecific amplification of regions of DNA which can be easily seen when the PCR product is analyzed by agarose gel electrophoresis (figure 2, lane 5). These nonspecific bands did not interfere with the GeneComb results because the probes are very specific for the desired product. A123 (DH5α) produces extra bands which are not detected by the comb (figure 1, tooth 6).

The GeneComb, unlike the agarose gel, was able to distinguish between SLT I and SLT II. The gel gives only one band which may indicate the presence of either toxin or both. On the comb, we utilized two probes MK1, in the lower position, and MK2, in the upper position, for the two SLT toxins. We were then able to detect the presence of the toxins separately. All of the agarose gel electrophoresis and GeneComb results are given in table 4.

The sensitivity of the GeneComb was compared to that of electrophoresis. The PCR product was diluted 1000-fold and easily detected by either method thus suggesting similar sensitivities. Published accounts (Reinhartz et. al., 1993) report a greater senstivity of the chromatographic assay.

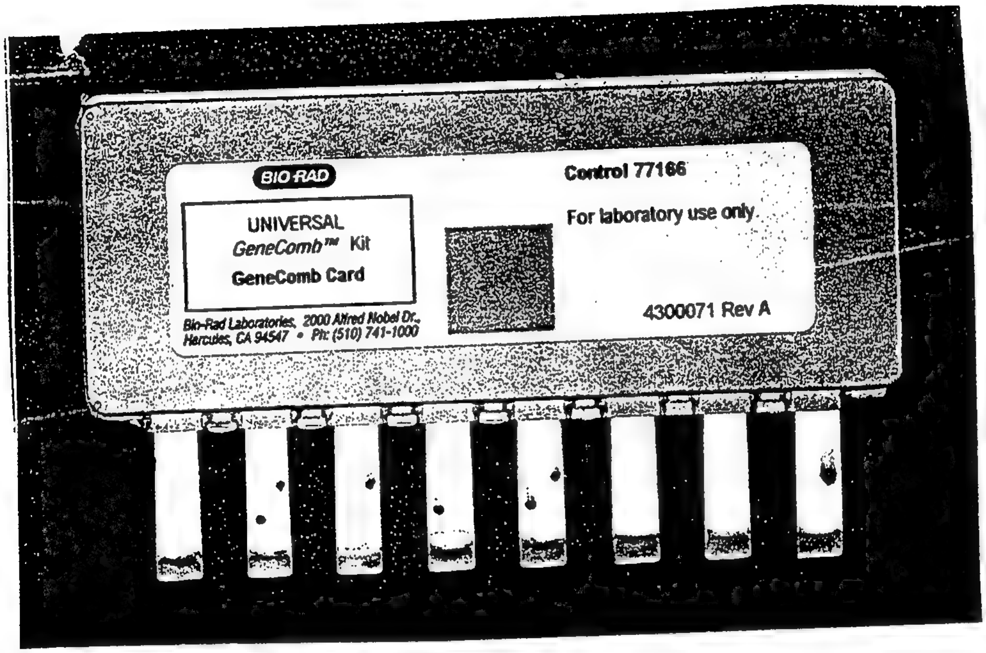


Figure 1: GeneComb.

The lower spot is a 1:1 mixture of the SLT I (MK1) and SLT II (MK2) probes and the upper spot is the 60-MDa plasmid probe (MFSPP).

Tooth 1=blank; tooth 2= A45; tooth 3= A57; tooth 4=A14; tooth 5=A45; tooth 6=A123 (DH5α); tooth 7=lysis; tooth 8=control

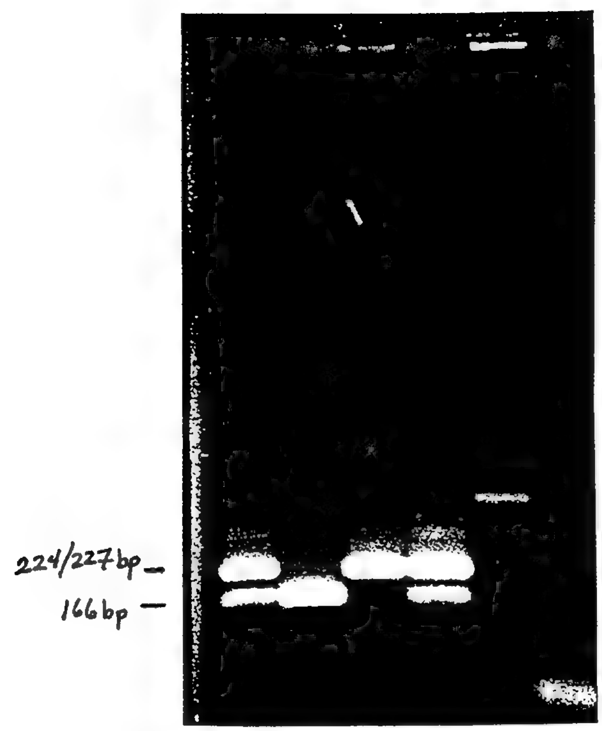


Figure 2: Agarose gel.

Lane 1=A45; lane 2=A57; lane 3=A14; lane 4=A45, lane 5=DH5α; and lane 6=lysis buffer

Table 4: Agarose gel electrophoresis and GeneComb results

ID #	Agarose Gel Electrophoresis		GeneComb Assay		
	Plasmid (166 bp)	SLT (224/227 bp)	Plasmid	SLT I	SLT II
A5	-	+	-	+	+
A11	+	+	+	+	+
A13	-	+	-	+	-
A14	-	+	-	-	+
A21	+	+	+	+	+
A40	-	-	-	-	-
A45	+	+	+	+	+
A57	+	+	+	+	-
A58	+	-	+	-	-
A59	-	+	-	+	+
A63	-	-	-	-	-
A98	+	+	+	+	+
A112	+	+			
A123	-	-	-	-	-
A124	+	+	+	+	+
A135	-	-	-	-	-
A136	-	-	-	-	-
A139	-	-	-	-	-

CONCLUSION

The detection of the plasmid, SLT I and SLT II toxin genes using the polymerase chain reaction could be a method of detecting the presence of *Escherichi coli* O157:H7 when other biochemical and serological assays are not practical. Using the Universal GeneComb from BioRad would further simplify the process by eliminating the need for agarose gel electrophoresis. The GeneComb assay is rapid, simple, and can be easily mastered.

The Gene Comb assay was very effective in detecting the plasmid, SLT I and SLT II amplified sequences. Within thirty minutes of the PCR amplification reaction, the resulting amplicons can be completely analyzed for the presence of the desired amplified sequences. This assay gives a considerable time savings as compared to agarose gel electrophoresis. Electrophoresis would involve casting, loading, staining, and visualization of the gel by UV light. Also in electrophoresis, errors may occur in intrepreting the results. Using the GeneComb, reading the results is as simple as noting the presence of a spot.

The major drawbacks associated with the comb are (1) the cost of the comb itself, $65.00 per comb, and (2) the additional costs of having the primer(s) biotinylated and having the probes synthesized. But given the time required for casting, loading, and staining a gel, these adverse factors associated with the comb are acceptable.

REFERENCES

Abbott, S.L., D. F. Hanson, T.D. Felland, S. Connell, A.H. Shum, and J. M. Janda. 1994. *Escherichia coli* O157:H7 generates a unique biochemical profile on MicroScan conventional gram-negative identification panels. Journal of Clinical Microbiology 32(3): 823-824.

Acheson, D.W.K, and G.T. Keusch. 1996. Which Shiga toxin-producing types of *E. coli* are important? ASM News 62(6): 302-306.

Anonymous. 1996. San Antonio Express News, July 27, 1996, page 12A.

Begum, D., N.A. Strockbine, E.G. Sowers, and M.P. Jackson. 1993. Evaluation of a technique for identification of Shiga-like toxin-producing *Escherichia coli* by using polymerase chain reaction and digoxigenin-labeled probes. Journal of Clinical Microbiology 31(12): 3153-3156.

Bettelheim, K.A., H. Evangelidis, J.L. Pearce, E. Sowers, and N.A. Strockbine. 1993. Isolation of a *Citrobacter freundii* strain which carries the *Escherichia coli* O157 antigen. Journal of Clinical Microbiology 31(3): 760-761.

Brian, M.J., M. Frosolono, B.E. Murray, A. Miranda, E.L. Loopez, H.F. Gomez, and T.G. Cleary. 1992. Polymerase chain reaction for diagnosis of enterohemorrhagic *Escherichia coli* infection and hemolytic-uremic syndrome. Journal of Clinical Microbiology 30(7): 1801-1806.

Cebula, T.A., W.L. Payne, and P. Feng. 1995. Simultaneous identification of strains of *Escherichia coli* serotype O157:H7 and their Shiga-like toxin type by mismatch amplification mutation assay-multiplex PCR. Journal of Clinical Microbiology 33(1): 248-250.

Fratamico, P., S. K. Sakitey, M. Wiedmann, and M.Y. Deng. 1995. Detection of *Escherichia coli* O157:H7 by multiplex PCR. Journal of Clinical Microbiology 33(8): 2188-2191.

Gherna, R., P. Pienta, R. Cote, eds. 1992. American Type Culture Collection Catalogue of Bacteria and Phages. 18th ed. ATCC, Rockville, MD.

Gunzer, F., H. Böhm, H. Rüssmann, M. Bitzan, S. Aleksic, and H. Karch. 1992. Molecular detection of sorbitol-fermenting *Escherichia coli* O157 in patients with hemolytic-uremic syndrome. Journal of Clinical Microbiology 30(7): 1807-1810.

Jackson, M.P., R.J. Neill, A.D. O'Brien, R.K. Holmes, and J.W. Newland. 1987. Nucleotide sequence analysis and comparison of the structural genes for Shiga-like toxin I and Shiga-like toxin II encoded by bacteriophages from *Escherichia coli* 933. FEMS Microbiology Letters 44: 109-114.

Karch, H., and T. Meyer. 1989. Single primer pair for amplifying segments of distinct Shiga-like toxin genes by polymerase chain reaction. Journal of Clinical Microbiology 27(12): 2751-2757.

Levine, M.M., J. Xu, J.B. Kaper, H. Lior, V. Prado, B. Tall, J. Nataro, H. Karch, and K. Wachsmuth. 1987. A DNA probe to identify enterohemorrhagic *Escherichia coli* of O157LH7 and other serotypes that cause hemorrhagic colitis and hemolytic uremic syndrome. The Journal of Infectious Diseases 156: 175-181.

Reinhartz, A., S. Alajem. A. Samson, and M. Herzberg. 1993. A novel rapid hybridization technique: paper chromatography hybridization assay (PACHA). Gene 136: 221-226.

Rice, E.W., E.G. Sowers, C.H. Johnson, M.E. Dunnigan, N.A. Strockbine, and S.C. Edberg. 1992. Serological cross-reactions between *Escherichia coli* O157 and other species of the genus *Escherichia*. Journal of Clinical Microbiology 30(5): 1315-1316.

Strockbine, N.A., L.R.M. Marques, J.W. Newland, H.W. Smith, R.K. Holmes, and A.D. O'Brien. 1986. Two toxin-converting phages from *Escherichia coli* O157:H7 strain 933 encode antigenically distinct toxins with similar biologic activities. Infection and Immunity 53(1): 135-140.

Effect of Dissolved Organic Matter on Fe(11) Trnsport in GroundwaterAircraft

Eric J. Henry
Graduate Research Assistant
Department of Civil and Environmental Engineering

Washington State University
Pullman, WA 99164

Final Report for:
Graduate Student Research Program
Armstrong Laboratory

Sponsored by:
Air force Office of Scientific Research
Bolling Air Force Base, DC

and

Armstrong Laboratory

September 1996

EFFECT OF DISSOLVED ORGANIC MATTER ON
Fe(II) TRANSPORT IN GROUNDWATER

Eric J. Henry
Graduate Student
Department of Civil and Environmental Engineering
Washington State University

Abstract

As the focus of groundwater remediation efforts shifts increasingly towards natural attenuation as an alternative method for subsurface restoration, a great deal of research must now focus on methods for documenting and quantifying such intrinsic remediation. One indicator of natural attenuation under iron-reducing conditions is concentration of dissolved Fe(II). However, if Fe(II) is to be used to quantify the degradation of groundwater contaminants the processes controlling Fe(II) transport in the subsurface must be better understood.

Dissolved metals, such as Fe(II) can interact with dissolved organic matter (DOM) to produce both mobile and immobile complexes. These complexes may display sorptive characteristics different than those of the dissolved metal alone, thus potentially facilitating or retarding transport of the metal. Microcosm sorption studies were conducted to determine the effects of DOM on Fe(II) sorption to aquifer solids from 3 U.S. Air Force Bases as a function of ionic strength (I). DOM at a concentration of 32 mg TOC/L resulted in a marked increase in the sorption of Fe(II) to each of the aquifer solids at $I = 0.01$ M, as judged by Freundlich non-linear isotherm fits of the data. Sorption of Fe(II) in the presence of DOM at $I = 0.1$ also increased over that of DOM-free systems but was less than that in the $I = 0.01$ systems, indicating a inverse relationship between Fe(II) sorption and ionic strength.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

EFFECT OF DISSOLVED ORGANIC MATTER ON
Fe(II) TRANSPORT IN GROUNDWATER

Eric J. Henry

Introduction

As the focus of groundwater remediation efforts shifts increasingly towards natural attenuation as an alternative method for subsurface restoration, a great deal of research must now focus on methods for documenting and quantifying such intrinsic remediation. One indicator of natural attenuation under iron-reducing conditions is concentration of dissolved Fe(II). During the microbial degradation of organic contaminants in anaerobic systems with Fe(III) the electron acceptor, Fe(III) is reduced to Fe(II), and Fe(II) concentrations increase (Wiedemeier *et al.*, 1995). Fe(II) can be used to quantify the degradation of groundwater contaminants if the processes controlling Fe(II) transport in the subsurface are understood.

Sorption has been identified as having a potentially significant effect on the transport of metals in the subsurface (Coughlin and Stone, 1995). Additionally, dissolved metals, such as Fe(II) can interact with dissolved organic matter (DOM) to produce both mobile and immobile complexes. These complexes may enhance or decrease sorption distribution between dissolved and solid phases, thus potentially facilitating or retarding transport of the metal. The ability of DOM, such as natural humic substances, to perform in this regard is dependent upon contaminant-humic, contaminant-sediment, and humic-sediment affinities as well as the kinetics of each of these interactions (Johnson and Amy, 1995). The strong affinity for metals to form complexes with humic substances has been well documented (Sposito, 1986; Perdue, 1989; Oden *et al.*, 1993; Benedetti *et al.*, 1995), and therefore a humic acid was chosen as a representative DOM for this investigation. Piana (1995) has described the sorption of Aldrich Humic Acid (AHA) onto aquifer solids from three U.S. Air Force Bases: Columbus AFB, Mississippi, Barkesdale AFB, Louisiana and Blytheville AFB, Arkansas, and Libelo (in prep.) has described the sorption of Fe(II) onto the same three aquifer solids. In order to supplement the data collected by Libelo and Piana, and maximize the relevancy of this research project, it was decided to investigate the effect of AHA on Fe(II) sorption onto the same aquifer solids.

Materials and Methods

General Description:

Batch sorption studies were conducted on three aquifer solids at two ionic strengths to describe the effect of natural organic matter on Fe(II) sorption. Dialysis tubing was used to achieve phase separation between the free ferrous ion and the ferrous-organic complex in order to describe the partitioning of Fe(II) to the humic acid. All experiments were carried out in 20 mL glass serum vials under anaerobic conditions to ensure that decreases in the aqueous Fe(II) concentration due to oxidation to Fe(III) were negligible. Anaerobic conditions were established and maintained either through bench-top nitrogen purging or operation within an anaerobic glove box. All solutions and dilution waters were nitrogen purged for 1 hour before use.

Aquifer Solids:

The substrates investigated in this study were obtained from Dr. T. Stauffer, Armstrong Laboratory, Tyndall AFB, FL, and consisted of unconsolidated sediments from aquifers at 3 Air Force Bases: Columbus AFB, Mississippi, Barkesdale AFB, Louisiana and Blytheville AFB, Arkansas. Specifically, the fraction of each sediment less than 2 millimeters was used in these experiments. Stauffer (1987) and Libelo (1995) have characterized each of the sediments and a summary is presented in Table 1.

Ionic Strength:

McCarthy and Zachara (1989) note that solution chemistry, including ionic strength and pH, can have an impact on contaminant sorption, and that such issues remain as areas which merit additional investigation. A number of researchers, including Westall *et al.* (1995), and Zachara *et al.* (1994) have shown metal sorption and complexation with humic acid to be a function of the geochemistry of the system. To determine the impact of ionic strength on Fe(II) sorption, all experiments described in this report were performed in ionic strength buffers of 0.01 and 0.1 M sodium perchlorate ($NaClO_4$) dissolved in Milli-Q ($\approx$18MΩ) water. In the interest of time and in order to keep the chemistry of the experimental systems as

simple as possible to avoid possible buffer ion effects, no controls on pH were instituted. The pH of the samples following equilibration was measured and ranged from 4.5-5.5.

Humic Acid Solutions and Analysis:

Aldrich Humic Acid (Aldrich, Milwaukee, WI) was used as a representative dissolved organic matter. 1000 mg/L humic acid solutions were prepared in ionic strength buffer solutions of 0.1 M and 0.01 M $NaClO_4$. Preliminary experiments showed the dissolved AHA to be approximately 32 % total organic carbon (TOC) by weight. All TOC measurements were made using a Shimadzu TOC-5000 combustion/non-disperse infrared gas analysis system (Shimadzu Corp., Kyoto, Japan). Kim *et al.*, (1990) provide further characterization of Aldrich HA which is summarized in Table 2.

Most groundwaters have dissolved organic carbon concentrations below 2 mg C/L, with a median value of about 0.7 mg C/L (Leenheer *et al.*, 1974, referenced by Drever, 1988). Aiken *et al.* (1985, referenced by Piana, 1995) report TOC values of natural waters between 1 and 30 mg C/L. A total organic carbon value at the high end of typical values, 32 mg C/L, was used for all experiments. It was expected that any variation in the sorption characteristics of Fe(II) due to the presence of DOM would increase with increasing TOC concentration and the use of a high concentration of TOC should therefore make those effects more pronounced and easily identified.

Fe(II) Solutions and Analysis:

Ferrous iron stock solutions were prepared by dissolving ferrous ammonium sulfate, $Fe(NH_4)_2(SO_4)_2$, in 0.1 and 0.01 M ionic strength buffers under nitrogen gas purge. Analysis of aqueous Fe(II) was conducted using a variation of the Ferrozine method described by Stookey (1970) and Gibbs (1976). Briefly, Ferrozine reagent forms a complex with Fe(II) which allows spectrophotometric analysis with the maximum absorbance occurring at a wavelength of 562 nanometers.

Ferrozine reagent was prepared by dissolving approximately 12 grams of HEPES buffer and 1 gram of Ferrozine (Aldrich) and diluting to 1 liter with Milli-Q water. Serial dilutions of the Fe(II) stock solutions were made, again using nitrogen purged ionic strength solutions, and a calibration curve developed. All

calibration standards, as well experimental samples, were prepared prior to analysis by the addition of 100 microliters of sample to 2 mL of Ferrozine reagent in a glass sample vial. Analyses were performed using a Cary 3E UV-Visible Spectrophotometer (Varian Instruments). The calibration curve was found to be linear throughout the range of interest for these experiments (0-100 mg Fe(II)/L).

Dialysis Tubing:

The molecular weights of dissolved organic matter range from 500 to 30,000 (Amy *et al.*, 1992). Particles within this size range are not settleable using centrifugation and therefore it was necessary to utilize another technique to obtain phase separation between ferrous ion in the free ion form, denoted Fe^{2+}, and that which is complexed with HA, denoted Fe^{2+}-HA. Dialysis tubing was chosen for phase separation because it is available in molecular weight cutoffs (MWCO) as low as 1000 and other researchers have shown it to be both effective for HA separation (Carter and Suffet, 1982; Zachara *et al.*, 1994) and durable within a system containing sediment (Allen-King *et al.*, 1995). Additionally, dialysis tubing has the advantage of achieving phase separation and equilibration simultaneously.

Spectra/Por 7, Regenerated Cellulose Membrane Dialysis Tubing, MWCO = 1000, flat width = 18 mm, was chosen for use because it is free of heavy metals and may be easily sealed by tying. The tubing is packaged in a dilute solution of sodium azide and required rinsing before use. The rinsing procedure consisted of placing the tubing in a 600 mL beaker filled with Milli-Q water, shaking the beaker for 1 minute, and then pouring the water into a waste container. This procedure was repeated 5 times. When not in use, rinsed tubing was stored in Milli-Q water.

Following rinsing, the dialysis tubing was cut into 16 centimeter lengths and one end tied. The tubing was then opened and 3 mL of ionic strength solution, appropriate to the given experiment, was pipetted into the tubing and the open end tied, forming a dialysis 'bag'. Excess tubing was trimmed away before placing the dialysis bag into the sample vials. An illustrative description of the experimental setup is shown in figure 1.

Rate Experiments:

Equilibrium is reached within 24 hours with regard to both HA sorption and Fe(II) sorption to each of the 3 substrates (Piana, 1995; Libelo, in prep.). Therefore, so that the experimental run time could be minimized, it was desired to show that the diffusion of Fe(II) through the bag reached equilibrium within 24 hours. The time required for diffusion equilibrium was determined by placing a dialysis bag filled with 5 mL of 0.1 M ionic strength solution into a sample vial containing 14 mL of 60 mg/L Fe(II) solution. Enough vials were prepared in this manner to allow sampling at 4 different times, in triplicate. At time increments of 1, 2, 4, and 7 days, samples were withdrawn from the inside and outside of the bag. Change in Fe(II) concentration within the bag was not detectable at times greater than 1 day, thus an equilibration period of 1 day was deemed to be satisfactory.

Analogous experiments were conducted to ensure that diffusion of AHA through the bag within the 1 day equilibration period was negligible. The TOC concentration within the bag after 1 day was approximately 5.5 % of the total TOC in the system. Comparison to controls indicated an increase with time in the total TOC present in the systems which contained dialysis bags even though the outer TOC concentration did not increase noticeably. This suggests that the TOC within the bag was derived from the dialysis tubing and AHA diffusion was deemed negligible.

Equilibrium Adsorption Isotherms:

Fe(II) sorption isotherms were developed at a constant TOC of 32 mg C/L, at ionic strengths of 0.01 and 0.1, in triplicate. Ferrous iron solutions were made at five different concentrations (approximately 75, 57, 38, 19, and 8 mg Fe(II)/L). Two grams of sediment, 1.4 mL of 1000 mg/L humic acid (320 mg C/L), and 12.6 mL of the appropriate ferrous iron solution were added to a 20 mL serum vial. A dialysis bag was then filled with 3 mL of ionic strength solution, placed in the vial, and the vial sealed. Controls were also prepared at the same dilutions, without sediment or humic acids.

Following 1 day equilibration on a rotary shaker, samples were removed and centrifuged (Damon/IEC EPR-6000) at 2000 rpm for 5 minutes. The solutions from inside and outside the dialysis bag

were then extracted using a volumetric syringe and 100 uL of each sample filtered (0.45 um) and added to 2 mL of Ferrozine for Fe(II) analysis. Following spectrophotometric analysis of all samples the total mass of Fe(II) added to each vial was determined using controls. The concentration of Fe(II) sorbed was then calculated using a mass balance and isotherms developed.

Sorption to solid substrates is often described as a linear partitioning of the contaminant between aqueous and solid phases:

$$C_s = K_d C_w \qquad (1)$$

where, Cs and Cw are the concentrations of the solute sorbed to the solid and dissolved in water, respectively, and Kd is the partitioning coefficient. Such a distribution assumes that sorption is independent of contaminant concentration and is defined as a linear isotherm. Frequently, however, a distribution of contaminant between the aqueous and solid phase which is non-linear with respect to the dissolved species is observed (Grathwohl, 1990). These non-linear partitioning distributions may be described using the Freundlich isotherm, equation 2:

$$Cs = K_f C_w^{1/n} \qquad (2)$$

where Cs and C_w are as defined as above, K_f is the Freundlich partitioning coefficient, and 1/n is a constant describing the dependence on solute concentration. The data from these experiments were fit with each type of isotherm and the results are presented in the following section.

Results And Discussion

At a TOC concentration of 32 mg/L, there was no statistical difference between the Fe(II) concentrations inside and outside of the dialysis bag. It was not possible, therefore, to describe the partitioning between of Fe(II) between the free ion phase, Fe^{2+}, and the humic complexed phase, Fe^{2+}-HA. The overall effect of AHA on Fe(II) sorption can be evaluated by the comparison of isotherms from systems containing AHA to isotherms from HA-free systems. The data from the HA-free systems have been kindly provided by Libelo (personal communication). In all systems (fixed HA concentration, fixed ionic strength) the order of decreasing affinity for Fe(II) sorption as a result of substrate type proceeds as Blytheville >

Columbus > Barkesdale. As an example, the isotherms for each substrate in an I = 0.01 system containing HA are given in figure 2.

The sorption isotherms depicted in figures 3 through 8 graphically illustrate the influence of HA on Fe(II) sorption at I = 0.01 and I = 0.1. The data were also modeled using linear and Freundlich isotherm expressions and the results are given in table 3. As judged by the r^2 values, the Freundlich isotherm typically yields the best fit and further discussion of the characteristic curves will be based on the Freundlich parameters.

Because both the K_f and 1/n values differ for each set of experiments it is difficult to compare the systems simply based on these values. As a semi-quantitative means of comparison, the sorbed equilibrium concentration was calculated according to the Freundlich equation corresponding to a dissolved Fe(II) concentration of 15 mg/L and these values are given in table 3. This concentration was chosen as an approximation of Fe(II) values occurring within an iron reducing environment of a polluted aquifer as reported by Albrechtsen and Christensen (1994).

From the graphs (3-8) it may be observed that the effect of AHA at low concentrations of Fe(II) is minimal, as should be expected according to both isotherm models. Of interest, however, is the trend which may be observed at increasing Fe(II) concentrations. Figures 3, 5, 7, and 8 reveal an marked increase in sorption of Fe(II) in the presence of AHA over that of HA-free systems. This conclusion is also supported by the calculated estimates of C_s in table 3. Some difficulty is associated with the use of figures 6 and 8 for supposition because the two isotherms being compared in each case encompass greatly varying Fe(II) concentrations.

Figures 9 through 11 contain all 4 isotherms (I = 0.1 and 0.01, 32 mg C/L and No AHA) for each substrate, and suggest that the sorption of Fe(II) in the presence of AHA decreases with increasing ionic strength. Other researchers have also shown this to be the case in metal-organic systems (Zachara *et al.*, 1994; Westall *et al.*, 1995) and associated this decrease in sorption with an decrease in the affinity of the organic for complexation with the dissolved metal at higher ionic strengths.

Conclusions and Recommendations for Future Work

The presence of dissolved organic matter has been shown to increase the sorption of dissolved Fe(II) to aquifer solids. This effect is enhanced as the ionic strength of the system decreases. The implication these results have for the use of dissolved Fe(II) as an indicator of natural attenuation is that measured Fe(II) concentrations in groundwater may underpredict the true production of Fe(II) during microbial degradation of organic contaminants under iron reducing conditions. Though this underprediction would yield a conservative estimate of the time necessary to achieve adequate remediation, such estimates are inefficient and may result in wasted time, money, and resources.

In order to obtain the best possible estimates of contaminant degradation rates additional research must be conducted to determine the effects of system geochemistry, including pH, redox, and the presence of competing sorbates on Fe(II) sorption in the presence of dissolved organic matter. Once such data is collected it should be possible to combine it with the data base already established by Piana (1995) and Libelo (in prep) to model Fe(II) sorption dynamics in the subsurface.

Acknowledgments

This work was funded by the Air Force Office of Scientific Research, Bolling AFB, Washington DC. I would like to sincerely thank the Armstrong Laboratory personnel, notably Dr. Laurence Libelo and Dr. Thomas Stauffer. Marlene Cantrell and Eila Burr of Applied Research Associates were always happy to provide laboratory assistance and Dr. Richelle Allen-King and Dr. William MacIntyre provided additional support during the course of this project.

Literature Cited

Aiken, G.R., McKnight, D.K., Wershaw, R.L. and P. MacCarthy, eds., Humic Substances in Soil, Sediment, and Water; Geochemistry, Isolation and Characterization. John Wiley and Sons, New York, 1985.

Albrechtsen, H. and T.H. Christensen, "Evidence for microbial iron reduction in a landfill leachate-polluted aquifer (Vejen, Denmark)," Appl. and Environ. Microbiol. 60(11) 3920-3925 (1994).

Allen-King, R.M., H. Groenevelt and D.M. Mackay, "Analytical method for the sorption of hydrophobic organic pollutants in Clay-Rich Materials," Environ. Sci. and Technol., 29(1): 148-153 (1995).

Amy, G.L., Sierka, R.A., Bedessem, J., Price, D. and L. Tan, "Molecular size distributions of dissolved organic matter," JAWWA, 84: 67-75.

Benedetti, M.F., Milne, C.J., Kinniburgh, D.G., Van Riemsdijk, W.H. and L.K. Koopal, "Metal ion binding to humic substances: Application of the non-ideal competitive adsorption model," Environ. Sci. and Technol., 29(2): 446-457 (1995).

Carter, C.W. and I.H. Suffet, "Binding of DDT to dissolved humic materials," Environ. Sci. and Technol., 16(11): 735-740 (1982).

Coughlin, B.R. and A.T. Stone, "Nonreversible adsorption of divalent metal ions (Mn(II), Co(II), Ni(II), Cu(II), and Pb(II)) onto goethite: Effects of acidification, Fe(II) addition, and picolinic acid addition," Environ. Sci. and Technol., 29(9): 2445-2455 (1995).

Drever, J.I., The Geochemistry of Natural Waters, 2nd ed.. Prentice Hall, New Jersey, 1988.

Gibbs, C.R., "Characterization and application of ferrozine iron reagent as a ferrous iron indicator," Anal. Chem., 48(8): 1197-1201 (1976).

Grathwohl, P., "Influence of organic matter from soils and sediments from various origins on the sorption of some chlorinated aliphatic hydrocarbons: Implications on Koc correlations," Environ. Sci. and Technol., 24: (1995).

Johnson, W.P. and G.L. Amy, "Facilitated transport and enhanced desorption of polycyclic aromatic hydrocarbons by natural organic matter in aquifer sediments," Environ. Sci. and Technol., 29(3): 807-817 (1995).

Kim, J.I., G. Backau, R. Klenze, D.S. Rhee and H. Wimmer, "Characterization and complexation of humic acid," Institut fur Radiochemie der Technischen Universitat Munchen-Final Report for the CEC project Mirage II, 1990.

Leenheer, J.A., "Fractionation techniques for aquatic humic substances," Humic Substances in Soil, Sediment, and Water, G.R. Aiken *et al.*, eds. Wiley-Interscience, New York, pp. 409-429, 1985.

Libelo, E.L., "Studies on processes controlling the input of agricultural chemicals in groundwater to surface waters," Ph.D. Dissertation, The College of William and Mary, 1995.

Libelo, E.L., Armstrong Laboratory, Tyndall Air Force Base, FL, personal communication, 1996.

McCarthy, J.F. and J.M. Zachara, "Subsurface transport of contaminants," Environ. Sci. and Technol., 23(5): 496-502 (1989).

Mills, W.B., S. Liu and F.K. Fong, "Literature review and model (COMET) for colloid/metals transport in porous media," Ground Water, 29: 199-208, (1991).
Oden, W.I., G.L. Amy and M. Conklin, "Subsurface interactions of humic substances with Cu(II) in saturated media," Environ. Sci. and Technol., 27(6): 1045-1051 (1993).

Perdue, E.M., "Effects of humic substances on metal speciation," pp 281-295 in Aquatic Humic Substances: Influence on Fate and Treatment of Pollutants, eds. I.H. Suffet and P. MacCarthy, American Chemical Society, Washington, D.C., 1989.

Piana, M.J., "The sorption of humic acids onto aquifer material solids," Final Report for AFOSR Summer Graduate Student Research Program, Armstrong Laboratory, August 1995.

Sposito, G., "Sorption of trace metals by humic materials in soils and natural waters," CRC Critical Reviews in Environmental Control, 16(2): 193-229, (1986).

Stauffer, T.B., "Sorption of nonpolar organics on minerals and aquifer materials," Ph.D. Dissertation, The College of William and Mary, 1987.

Stookey, L.L., "Ferrozine-A New Spectrophotometric Reagent for Iron," Anal. Chem., 42(7): 779-781 (1970).

Westall, J.C., Jones, J.D., Turner, G.D. and J.M. Zachara, " Models for association of metal ions with heterogeneous environmental sorbents. 1. Complexation of Co(II) by Leonardite Humic Acid as a function of pH and NaClO4 concentration," Environ. Sci. and Technol., 29(4): 951-959 (1995).

Wiedemeier, T.H., Wilson, J.T., Kampbell, D.H., Miller, R.N., and J.E. Hansen, Technical Protocol for Implementing Intrinsic Remediation with Long-Term Monitoring for Natural Attenuation of Fuel Contamination Dissolved in Groundwater. Air Force Center for Environmental Excellence, Brooks Air Force Base, Texas, 1995.

Zachara, J.M., C.T. Resch and S.C. Smith, "Influence of humic substances on cobalt(II) sorption by a subsurface mineral separate and its minerologic components," Geochimica et Cosmochimica Acta, Vol. 58(2): 553-566 (1994).

Table 1. Characteristics of Aquifer Sediments

Aquifer	Total Organic Carbon %	Surface Area m^2/g	Dithionate Extractable Iron g/kg	% Sand	% Silt	% Clay
Barkesdale	0.0338	0.88	0.112	96.1	1.94	1.97
Blytheville	0.0676	9.32	1.640	95.6	3.06	1.34
Columbus	0.0596	5.78	0.914	70.8	17.50	11.68

Table 2. Elemental Composition and Physical Properties of Aldrich Humic Acid

C	H	N	S	O	Other	H/C ratio	O/C ratio	Aromatic C (110-165)	Aliphatic C (0-90)	Carboxyl C (165-190) PPM
41.72	4.37	0.25	1.90	36.93	14.83	1.03	0.51	40	41	14

Table 3. Linear and Freundlich Isotherm Parameters

Sediment	Condition	log K_f[a]	1/n[a]	r^2	C_s[b]	Kd[c]	r^2	n[d]
Barkesdale	No HA, I=.01	1.084 (1.052-1.116)	0.512 (.476-.549)	0.985	48.6	2.52	0.93	16
	HA, I=.01	1.182 (1.132-1.233)	0.646 (.603-.690)	0.987	87.4	3.58	0.94	15
	No HA, I=.1	1.350 (1.212-1.488)	0.336 (.279-.402)	0.903	55.6	0.26	0.76	15
	HA, I=.1	1.182 (1.011-1.354)	0.431 (.289-.573)	0.860	48.9	1.39	0.75	10
Blytheville	No HA, I=.01	1.693 (1.591-1796)	0.207 (.033-.381)	0.530	86.4	5.53	0.76	9
	HA, I=.01	1.734 (1.680-1.788)	0.539 (.476-.602)	0.963	233.3	10.61	0.93	15
	No HA, I=.1	1.722 (1.622-1.822)	0.307 (.260-.354)	0.992	121.1	0.30	0.74	18
	HA, I=.1	1.550 (1.411-1.689)	0.536 (.395-.676)	0.906	151.5	8.05	0.85	10
Columbus	No HA, I=.01	1.124 (1.101-1.180)	0.590 (.564-.617)	0.993	65.8	4.32	0.95	19
	HA, I=.01	1.516 (1.485-1.548)	0.502 (.471-.533)	0.990	127.8	4.52	0.91	15
	No HA, I=.1	1.244 (.942-1.545)	0.336 (.204-.468)	0.646	43.6	0.65	0.25	18
	HA, I=.1	0.957 (.915-.999)	0.646 (.612-.681)	0.992	52.1	2.08	0.97	15

a.) 95% confidence interval in brackets; b.) sorbed concentration, C_s, at C_w = 15 mg Fe(II)/L calculated using Freundlich isotherm coefficients; c.) linear isotherm not forced through zero; d.) n =number of samples

Figure 1. Schematic Drawing of Equilibrium Sorption Experiment Setup

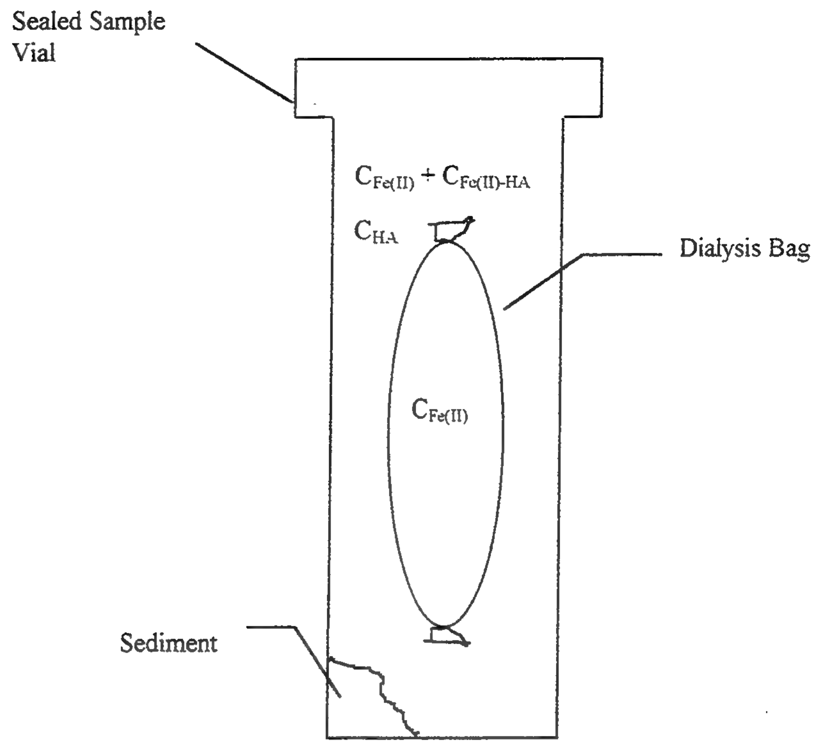


Figure 2. Example Fe(II) Sorption Isotherm
TOC = 32 mg/L, I = 0.01

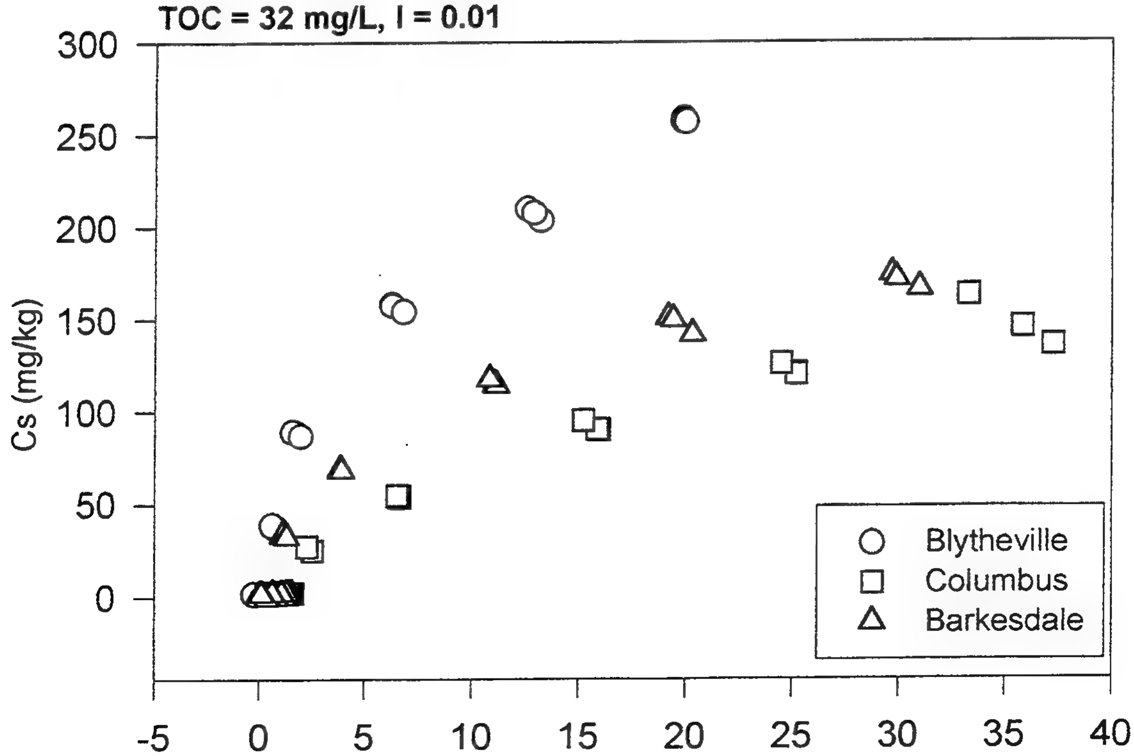


Figure 3. Freundlich Isotherm
Barkesdale, I = 0.01

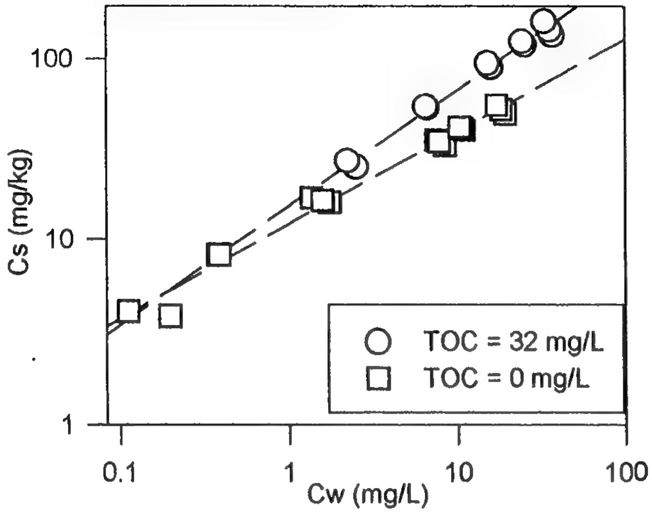


Figure 4. Freundlich Isotherm
Barkesdale, I = 0.1

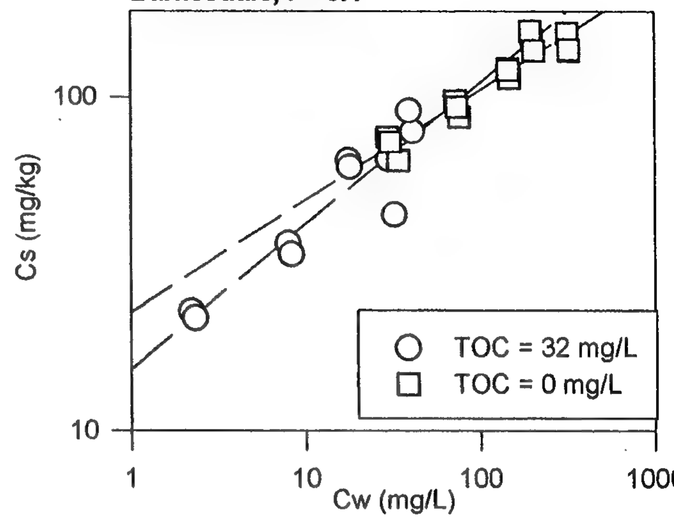


Figure 5. Freundlich Isotherm
Blytheville, I = 0.01

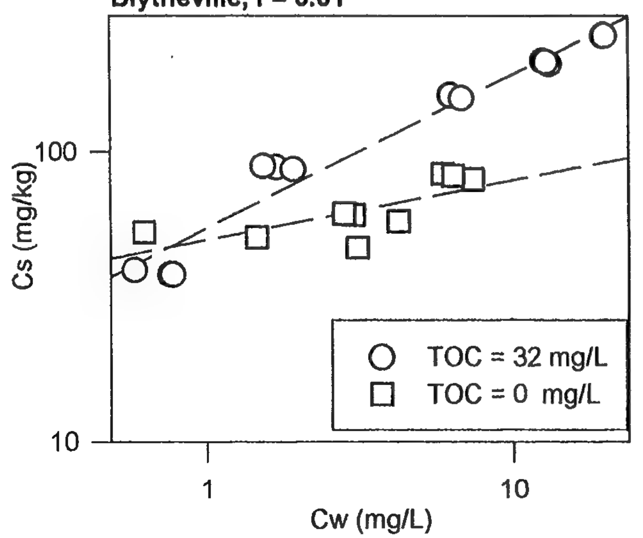


Figure 6. Freundlich Isotherm
Blytheville, I = 0.1

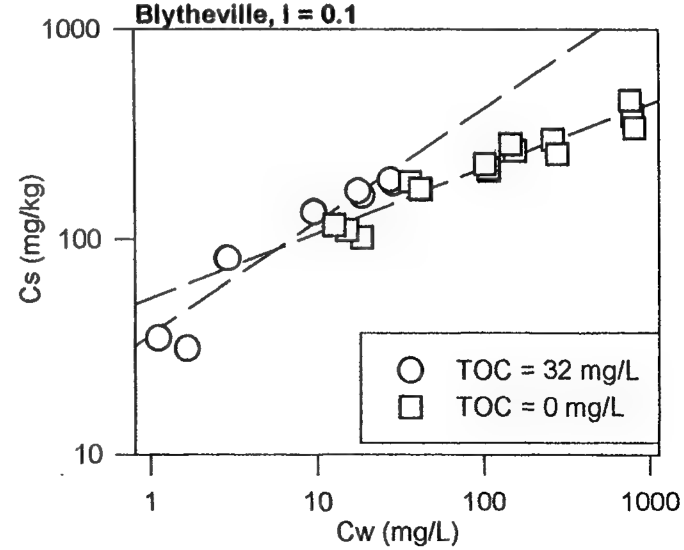


Figure 7. Freundlich Isotherm
Columbus, I = 0.01

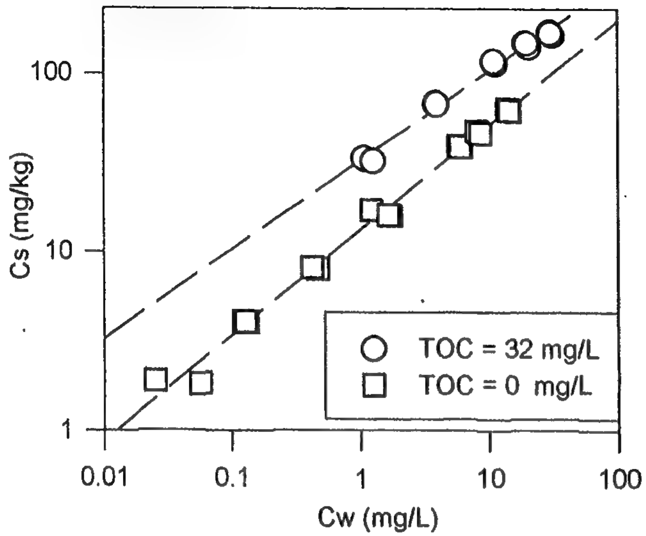


Figure 8. Freundlich Isotherm
Columbus, I = 0.1

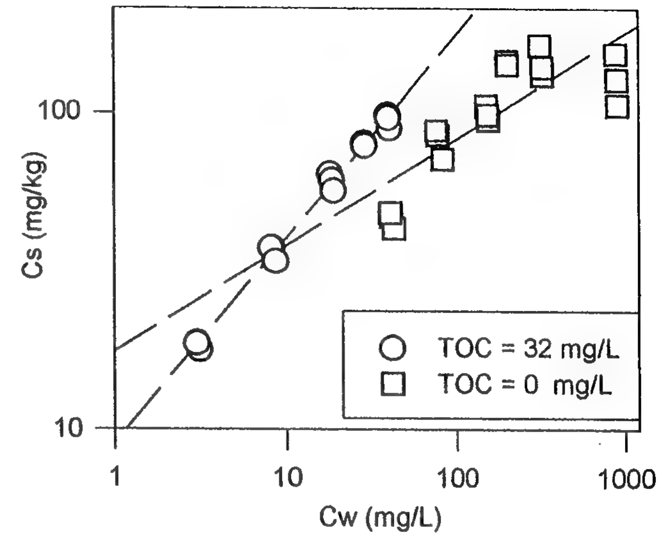


Figure 9. Sorption Isotherms: Barkesdale

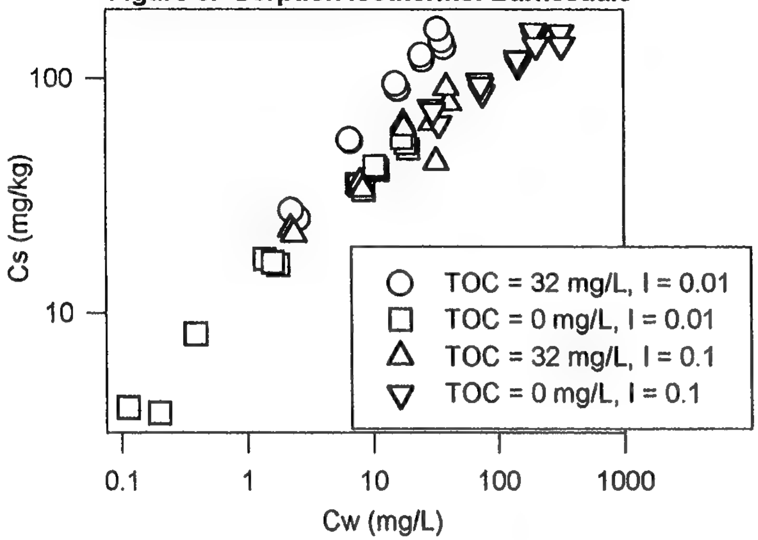


Figure 10. Sorption Isotherms: Blytheville

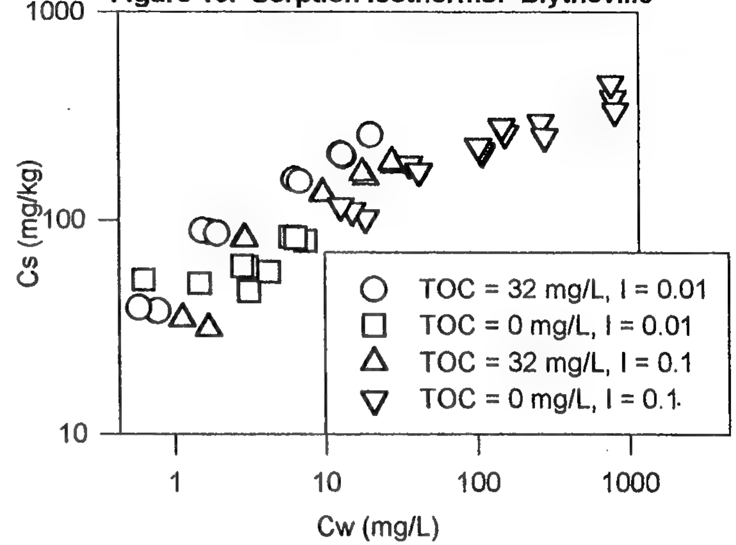
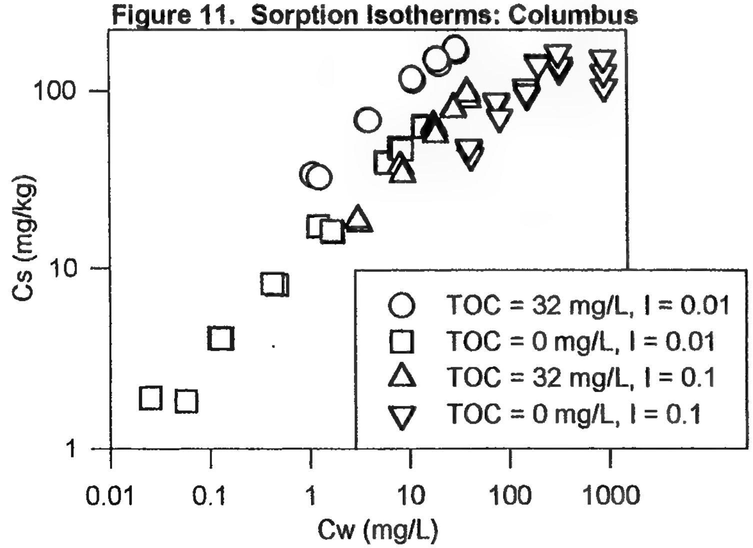


Figure 11. Sorption Isotherms: Columbus



Validity of ASVAB Selector AI and FSG for ASVAB and Paper and Pencil Forms 15, 16, and 17 and CAT Forms 1 and 2

David Herst
Graduate Research Assistant
Dept of Psychology

University of South Flordia
4202 East Fowler Avenue, BEH 329
Tampa, Florida 33620-8200

Final Report for:
Graduate Student Research Program
Armstrong Laboratory

Sponsored by:
Air force Office of Scientific Research
Bolling Air Force Base, DC

and

Armstrong Laboratory

August 1996

Validity of ASVAB Selector AI and FSG for ASVAB Paper and Pencil Forms 15, 16, and 17 and CAT Forms 1 and 2

David Herst
Graduate Student
Department of Psychology

Abstract

Validity of the Armed Services Vocational Aptitude Battery (ASVAB) forms 15, 16, 17 and Computer Adaptive Test (CAT) forms 1 and 2 was investigated. The Air Force constructs four vocational classification composites from the ten subtests within the ASVAB. They include Mechanical (M), Administrative (A), General (G), and Electronic (E). A proposed replacement for the A composite, A1, was also computed and investigated. Composite scores and final school grades (FSG) were compared for 44,929 non-prior enlisted military personnel in 110 technical schools using a two-tailed Pearson Product-Moment correlation. Average validities across classification composites, as well as validities for technical schools within each composite, were computed. Correlations were then corrected for range restriction.

All five of the composites showed significant average correlations with final school grades. Electronic composite scores had the highest average validity at .44 uncorrected, .70 corrected. This was followed by General at .34 (.47 corrected), Proposed A1 at .30 (.36 corrected), Mechanical at .28 (.46 corrected) and Administrative at .07 (.28 corrected). Of the 110 schools assessed, 6 had nonsignificant correlations between final school grades and ASVAB composite score. However, by using the Proposed A1 composite in place of the Administrative composite, that number was reduced to 4 and included a dramatic increase in predictive validity in assessing Administrative schools.

Introduction

In November 1948, the first airmen classification battery (AC-1A), was implemented, consisting of 13 test batteries and 8 composite scores known as aptitude indices, or AI (Weeks, Mullins, Vitola, 1975). By 1966 the Assistant Secretary of Defense Manpower and Reserve Affairs called for a uniform selection/classification testing battery to be used by all branches of the military. This became known as the Armed Services Vocational Aptitude Battery (ASVAB), which consists of four composite scores computed from ten subtests in the battery (Vitola, Mullins, Croll, 1973). The composite scores are: Mechanical (M), Administrative (A), General (G) and Electronic (E). New forms of the ASVAB have been developed approximately every four years, with six content-equivalent tests in each release. In 1980 the ASVAB was normed on a new nation-wide representative sample of youth (Military Testing Program Overview, 1994).

Changes in the type and skill requirements of Air Force jobs has prompted continuous evaluation of ASVAB composites as valid predictors of performance. ASVAB validity testing has centered on the use of FSG as a criterion, and the individual's ASVAB selector AI score for a specific AFSC, as the predictor. For example, Valentine (1977) compared nonprior service enlistees' scores on ASVAB-3 aptitude indices and their educational background to FSG. Valentine found that ASVAB scores alone were more accurate in prediction of FSG than educational background alone. Furthermore, when ASVAB scores were added to an equation which already included educational background, the test data provided a larger increase in prediction accuracy than if educational data were added to an equation with ASVAB scores. Finally, educational background was found to be more subject to bias against minorities and women than ASVAB test data.

Mullins, Earles, and Ree (1981) weighted aptitude components based on differences in technical school difficulty. Their results indicated that while weighting aptitude components substantially improved predictive validity, the inclusion of educational background did not add to this

increase. Ree and Earles (1992) investigated ASVAB forms 11, 12 and 13 and found the Electronics AI to have the highest predictive validity, regardless of the type of school being examined.

The purpose of the current study was twofold. First was to assess the validity of paper and pencil ASVAB forms 15, 16 and 17, as well as computer adaptive ASVAB forms 1 and 2, against FSG in technical schools. Like all previous ASVAB validity studies, it is assumed that the separate forms are "content and topologically equivalent" (Ree and Earles, 1992, p. 3). The purpose of the second portion of this study was to assess the validity of a new composite to be used in place of the current Administrative AI, the Proposed A1 composite. It is hypothesized that ASVAB aptitude index scores will be highly correlated with FSG, and that predictive validities between Proposed A1 and FSG will be greater than those of the current Administrative AI and FSG.

Method

Participants

Participants were 44,929 non-prior enlisted military personnel who had successfully completed technical school. All individuals accessed into the Air Force between 1 January 1989 and 1 October 1993 on Forms 15, 16, or 17 of the paper and pencil ASVAB, or on CAT-ASVAB Forms 1 or 2. The sample was predominately male (79.4%) and Caucasian (81%). The participants ranged in age from 17 to 27, with a modal age of 18. 99.7% had at least high school diplomas or GED. Sample characteristics are summarized in Table 1.

Instruments

Predictors. The ASVAB composites used for job assignment were the predictors for this study. The ASVAB is a multiple choice, multiple aptitude battery composed of 10 subtests. Factor analyses have shown that these tests measure verbal and quantitative ability, cognitive speed and technical knowledge (Military Testing Program Overview, 1994). The Air Force constructs four classification composites from the ten subtests: Mechanical (M),

Administrative (A), General (G), and Electronic (E). A proposed replacement for the Administrative composite, A1, was also included. These composites, also referred to as aptitude indexes (AI), are used to classify individuals into jobs. ASVAB composites are available as sum of standard scores or as percentiles. Sum of standard scores were used in this study. Both paper and pencil and CAT-ASVAB scores were included. ASVAB subtests and composite composition are shown in Tables 2 and 3.

Criterion. Final school grade (FSG) from technical training school was the criterion. FSG is the average of the multiple choice test grades obtained during technical school, and range between 70 and 99.

Procedure

Archival data obtained from the master files maintained by the Support Branch of the Manpower and Personnel Division of the Armstrong Laboratory of the Human Resource Directorate (AL/HRMX), were used in this study. The ASVAB scores of record, that is the scores on which the individual assessed, were used regardless of the number of times the individual has tested on the ASVAB. Technical schools were included that met three criteria. First, grades had been given in a numeric format. Second, a minimum of 100 individuals successfully completed the course. Finally, each school had only one selector AI.

Analysis Strategy

Within each school, uncorrected and corrected correlations were computed between the selector AI score and FSG. A Two-tailed Pearson Product-Moment correlation with an alpha of .05 was used to determine ASVAB AI validity. Correlations were corrected for range restriction using scores from 333,559 individuals who had applied to the Air Force while ASVAB forms 15, 16, 17 and CAT-ASVAB 1 and 2 were in use. For each AFSC, means, standard deviations and coefficients of determination where also calculated. Coefficients of determination were included to indicate the percentage of variance in FSG

accounted for by ASVAB selector AI score. Average corrected and uncorrected correlations, means, standard deviations and coefficients of determination were also computed for each subsample. Means and standard deviations for subtest and composite scores were computed for both the restricted and unrestricted sample. Finally, the proposed predictor A1 was used in place of AI A at Administrative AFSCs.

Results

Means and standard deviations for subtests and composites of the restricted and unrestricted samples are listed in Table 4 and Table 5. Table 6 presents the corrected and uncorrected interrcorrelation matrix between subtests for the entire sample. Only the correlation between Word Knowledge (WK) and Auto and Shop Information (AS) failed to be significant, $\underline{r}(44,929)=0.00$, $\underline{p}>.05$. Table 7 shows corrected and uncorrected intercorrelation matrix for composite scores. All correlations were significant. Corrected correlations were not computed for the A1 composite. Average validities for each aptitude area are presented in Table 8. Electronic AI had the highest correlation with final school grades at .44 ($\underline{N}=6527$, $\underline{p}<.05$), while Administrative AI scores indicated the lowest predictive validity at .07 ($\underline{N}=5857$, $\underline{p}<.05$). The removal of Numerical Operations and Coding Speed, and addition of Mathematical Knowledge subtest to the Administrative composite calculation increased average predictive validity to .30, ($\underline{p}<05$). The largest criterion mean was found in Electronic schools at 89.19, while the lowest appeared in General technical schools at 85.18.

Mechanical Aptitude Index. Mechanical technical schools in the Air Force include Tactical Aircraft Maintenance Apprentice, Helicopter Maintenance Apprentice and Utilities Systems Apprentice. Investigation of the 29 ($\underline{N}=11,690$) schools in this subsample indicated correlations between FSG and ASVAB Mechanical AI ranging from .11 ($\underline{N}=734$, $\underline{p}<.05$) to .47 ($\underline{N}=318$, $\underline{p}<.05$), which increased to .18 and .68 when corrected for range restriction. Thus, between 1% and 22% of the variance in predicting FSG is accounted for from Mechanical selector AI scores, as indicated by the coefficients of

determination. Criterion means for each AFSC in the sample were large, falling between a low of 83.24 and a high of 95.26.

Of the 29 Mechanical technical schools in the study, only the school for Survival Equipment Apprentice (AFSC 45833) had a predictive validity which was not significantly different from zero, $\underline{r}(226)=.12$, $\underline{p}>.05$. Results and descriptive statistics by school are listed in Table 9.

Administrative Aptitude Index. Only ten ($\underline{N}$=5,857) Administrative schools met the three-part criteria for use in this study. They ranged from Operations Resource Management Apprentice, to Traffic Management Apprentice, to Information Management Apprentice. Predictive validities were low, ranging between .01 ($\underline{N}$=326, $\underline{p}>.05$) and .29 ($\underline{N}$=108, $\underline{p}<.05$), .11 to .54 when corrected for range restriction. Variance in predicting FSG is considerably lower when using the Administrative AI, accounted for between 1% and 8% in this subsample. Criterion means fell between 81.40 and 90.30. Correlations between ASVAB Administrative AI scores and FSG for Traffic Management Apprentice schools 60230 [$\underline{r}(326)=.01$, $\underline{p}>.05$] and 60231 [$\underline{r}(348)=.08$, $\underline{p}>05$] were not significantly different than zero. A breakdown of descriptive statistics and validities by AFSC for Administrative schools can be seen in Table 10.

Proposed A1 Aptitude Index. The Proposed A1 composite was correlated with subject FSG from the subsample of ten Administrative technical schools. Predictive validity was found to increase for all schools, with the uncorrected range of correlations falling between .20 ($\underline{N}$=326, $\underline{p}<.05$) and .44 ($\underline{N}$=246 and $\underline{N}$=1151, $\underline{p}<.05$), corrected validities ranging between .28 and .63. As with the uncorrected validities, variance accounted for increased over the use of Administrative AI, ranging between 4% and 19% when using the Proposed A1 composite. All correlations were found to be significant. Descriptive statistics and validities by AFSC are listed in Table 11.

General Aptitude Index. General technical schools in the Air Force include a wide range of specialties from Electronic Signals Intelligence Exploitation Apprentice and Maintenance Data Systems Analysis to Radiological Apprentice and Surgical Service Apprentice. A total of 41 (N=20,855) General technical schools met the criteria to be included in this study. Uncorrected validities ranged between .04 (N=192, p>.05) and .52 (N=103, p<.05), .12 and .68 when corrected for range restriction due to rigorous selection procedures. Accordingly, variance accounted for ranged between a low of 0% and a high of 26%. Criterion means fell between a low of 83.76 and a high of 93.03. Predictive validities for Signals Intelligence Analysis Apprentice (AFSC 20230) [r(221)=.11, p>.05], AFSC 20834G [r(104)=.19, p>.05] and Maintenance Scheduling Apprentice (AFSC 39230) [r(192)=.04, p>.05] were not significant. Results by individual AFSC are presented in Table 12.

Electronic Aptitude Index. Air Force Electronic schools included training from Space Systems Operations Apprentice and Telephone Switching Apprentice to Biomedical Equipment Apprentice. Exactly 30 (N=6,527) technical schools met the criteria to be included in this subsample. Uncorrected predictive validities ranged between .25 (N=104 and N=111, p<.05) and .62 (N=103, p<.05), .81 and .47 when corrected for range restriction due to selection procedures. Coefficients of determination indicated that variance in predicting FSG from Electronic selector AI ranged between 5% and 38%, depending on the AFSC being tested. Criterion means were on average larger than other samples in the study, falling between a low of 85.52 and a high of 93.39. All predictive validities were significantly different from zero. Descriptive statistics and validities for each AFSC are listed in Table 13.

Discussion

Of the 110 technical training schools in this sample, only 6 were found to have predictive validities which were not significant. This number was reduced to 4 when the Proposed A1 composite was used in place of Administrative AI scores for Administrative AFSC. That leaves between 104 and 106 significant correlations, with the possibility of up to 6 resulting from Type I error. The small number of nonsignificant correlations demonstrates the validity of the ASVAB as classification tool. However, the reasons for the nonsignificant validities are unclear. Since all AFSCs contained at least 100 subjects, power can be eliminated as a possible factor. One explanation may lie in the nature of the AFSC being validated. Schools such as Traffic Management Apprentice require a wide range of skills such as preparing transportation requests, quality control documents, use of forklifts and checks of airline routings. These skills may not have been measured well by the ASVAB, resulting in a comparatively low Administrative AI score. Yet, placed in an environment which requires extensive hands-on training, an individual may thrive and thus score extremely well on tests given throughout the course of study. Hence the nonsignificant correlation between ASVAB AI and FSG. Additional research is recommended to determine the reasons for nonsignificant validities in select technical training schools.

The impressive number of significant correlations can be partially attributed to the large sample sizes in several of the schools studied. By considering a large sample to be 400 or more subjects, a full 26 technical schools (25%) could have significant correlations simply due to excessive power by way of sample size. However, using the conservative estimate that an $\underline{r}>.25$ and $\underline{N}=100$ is significant, 20 of the 26 schools indicated can be eliminated from consideration. Thus it can also be assumed that excessive power due to sample size is a factor in only 6% of the technical schools found to have significant validities. This is a strong indication that while correlations between ASVAB AI and FSG vary according to the type of AFSC, the aptitude indexes for ASVAB

forms 15, 16, 17 and CAT forms 1 and 2 are nevertheless valid in their predictions of performance.

The repeated revision and validity testing of the ASVAB since its inception has provided a better understanding of how to properly classify individual recruits. However, validation must remain a high priority in order to strive for better classification systems. This study investigated the validity of aptitude indexes in their prediction of FSG. Yet the validation of the subtest scores which comprise the aptitude indexes also need to be investigated at regular intervals, as should the ASVAB's impact on minority groups and women. Since subtest scores are used to create the ASVAB selector AI, their reliability and validity are crucial to continued fair and effective classification of recruits. These issues, while beyond the scope of this paper, represent a sample of the reasons for continual testing and revision of the ASVAB classification system in order to both increase predictive validity and insure accuracy across all groups.

References

Military Testing Program Overview: For New Members of the Defense Advisory Committee on Military Personnel Testing. Orientation material for new DAC members, 1994.

Mullins, C.J., Earles, J.A., and Ree, M. (1981). *Weighting of Aptitude Components Based on Differences in Technical School Difficulty* (AFHRL-TR-81-19, AD-a102 045). Brooks AFB, Tx: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Ree, Malcolm J. and Earles, James A. (1992). *Subtest and Composite Validity of ASVAB Forms 11, 12, and 13 for Technical Training Courses* (AL-TR-1991-0107, AD-A247 728). Brooks AFB, Tx: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Valentine, Jr., Lonnie D. (1977). *Prediction of Air Force Technical Training Success from ASVAB and Educational Background* (AFHRL-TR-77-18, AD-a041 0735). Brooks AFB, Tx: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Vitola, Bart M., Mullins, Cecil J., Croll, Sgt. Paul R. (1973). *Validity of Armed Services Vocational Aptitude Battery, Form 1, to Predict Technical School Success* (AFHRL-TR-73-7, AD-767 578). Brooks AFB, Tx: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Weeks, Joseph L., Mullins, Cecil J., Vitola, Bart M. (1975). *Airmen Classification Batteries from 1948 to 1975: A Review and Evaluation* (AFHRL-TR-75-78, AD-a026 470). Brooks AFB, Tx: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Table 1.
Participant Demographics.

Gender	Percent	Age	Percent
Male	79.4	17-18	29.2
Female	20.6	19-20	41.6
		21-22	17.4
		23+	11.8
Ethnicity	**Percent**	**Education**	**Percent**
White	81.0	No HS Diploma	0.1
Black	12.1	GED Diploma	0.8
Hispanic	3.9	HS Diploma	75.7
Am. Indian	0.3	Some College	18.8
Asian	1.8	Associates Degree	2.7
		Advanced Degree	1.7

Table 2.
ASVAB Subtests for Forms 15, 16, and 17 and CAT 1 and 2.

Subtest	Type of Tests
General Science (GS)	Power
Arithmetic Reasoning (AR)	Power
Word Knowledge (WK)	Power
Paragraph Comprehension (PC)	Power
Auto and Shop Information (AS)	Power
Mathematical Knowledge (MK)	Power
Mechanical Comprehension (MC)	Power
Electronics Information (EI)	Power
Numerical Operations (NO)	Speed
Coding Speed (CS)	Speed
Verbal (VE)**	N/A

*VE = WK + PC

Table 3.
ASVAB Composites for Forms 15, 16, and 17 and CAT 1 and 2.

Composite	Composition
Armed Forces Qual. Test (AFQT)	2(VE)+AR+MK
Mechanical (M)	MC+GS+2AS
Administrative (A)	WK+PC+NO+CS
Proposed Admin. (A1)	WK+PC+MK
General (G)	WK+PC+AR
Electronic (E)	GS+AR+MK+EI

Table 4.
Means and Standard Deviations for ASVAB Standardized Subtests and Sum of Standardized Composite Scores for the Restricted and Unrestricted Samples.

Variable	Predictor Type	Restricted Sample Mean	Restricted Sample S.D.	Unrestricted Sample Mean	Unrestricted Sample S.D.
Mechanical	Composite	215.79	25.36	207.63	29.08
Administrative	Composite	167.15	11.95	163.15	14.87
General	Composite	110.66	8.18	106.24	11.66
Electronic	Composite	219.48	20.01	210.26	24.95
GS	Subtest	54.56	6.56	52.32	7.72
AR	Subtest	55.46	6.14	52.76	7.69
WK	Subtest	54.96	4.22	53.28	5.81
PC	Subtest	55.19	4.34	53.54	6.03
NO	Subtest	55.96	5.84	54.72	6.75
CS	Subtest	55.98	6.74	54.94	7.22
AS	Subtest	52.85	8.28	51.10	8.93
MK	Subtest	56.69	7.04	54.13	7.98
MC	Subtest	55.54	7.56	53.10	8.51
EI	Subtest	52.78	7.96	51.05	8.59
VE	Subtest	55.21	3.89	53.48	5.59

Note. All calculations performed using N=44,929

Table 5.
Restricted Sample ASVAB Composite Percentile Means and Standard Deviations.

Composite	N	Mean	S.D.
AFQT	44,929	68.25	15.68
Mechanical	44,929	63.63	21.06
Administrative	44,929	71.22	18.30
General	44,929	67.21	16.35
Electronic	44,929	67.06	16.57

Table 6.
Uncorrected and Corrected Intercorrelation Matrix for ASVAB Subtest Standardized Scores.

	GS	AR	WK	PC	NO	CS	AS	MK	MC	EI	VE
VE	0.52**	0.30**	0.95**	0.69**	0.04**	0.07**	0.14**	0.30**	0.26**	0.27**	1.00
EI	0.50**	0.27**	0.27**	0.15**	-0.14**	-0.15**	0.61**	0.19**	0.53**	1.00	0.41
MC	0.48**	0.41**	0.25**	0.16**	-0.10**	-0.11**	0.54**	0.29**	1.00	0.60	0.44
MK	0.34**	0.56**	0.27**	0.24**	0.28**	0.20**	0.00	1.00	0.40	0.29	0.44
AS	0.39**	0.21**	0.15**	0.06**	-0.21**	-0.21**	1.00	0.13	0.60	0.66	0.31
CS	-0.09**	0.15**	0.04**	0.10**	.055**	1.00	-0.08	0.30	0.04	-0.01	0.22
NO	-0.07**	0.22**	0.01**	0.08**	1.00	0.59	-0.07	0.40	0.07	0.01	0.23
PC	0.31**	0.23**	0.45**	1.00	0.25	0.24	0.22	0.40	0.35	0.31	0.81
WK	0.51**	0.27**	1.00	0.64	0.19	0.19	0.32	0.41	0.42	0.41	0.97
AR	0.34**	1.00	0.50	0.46	0.35	0.26	0.35	0.64	0.54	0.41	.053
GS	1.00	0.52	0.64	0.49	0.12	0.08	0.48	0.46	0.58	0.57	0.65

Note. Uncorrected correlations in upper left of table, corrected correlations in lower right
**Correlations significant at $p<.01$, $N=44,929$

Table 7.
Uncorrected and Corrected Intercorrelation Matrix for ASVAB Aptitude Indexes.

	Mechanical	Administrative	General	Electronic	Proposed A1
Proposed A1	0.27**	0.44**	0.74**	0.76**	1.00
Electrical	0.66**	0.21**	0.77**	1.00	N/A
General	0.40**	0.40**	1.00	0.85	N/A
Administrative	-0.10**	1.00	0.60	0.46	N/A
Mechanical	1.00	0.18	0.57	0.74	N/A

Note. Uncorrected correlations upper left of table, corrected correlations in lower right
**Correlations significant at $p<.01$, $N=44{,}929$

Table 8.
Descriptive Statistics, Raw Validities and Validities Corrected for Range Restriction for ASVAB Selector AI and FSG.

			Predictor		Criterion				
AI	School	N	Mean	S.D.	Mean	S.D.	r	Cor.r[a]	Aj.r[2b]
Mechanical	29	11690	227.85	17.84	86.93	6.70	0.28**	0.46	0.08
Administrative	10	5857	169.95	8.88	86.14	6.13	0.07**	0.28	0.00
Proposed A1	10	5857	109.69	8.21	86.14	6.13	0.30**	0.36	0.09
General	41	20855	111.46	7.96	85.18	7.40	0.34**	0.47	0.12
Electronic	30	6527	236.09	13.02	89.19	4.87	0.44**	0.70	0.19

**Correlation significant at $p<.01$
[a]Correlation corrected for range restriction
[b]Coefficient of determination adjusted for population estimate

Table 9.
Descriptive Statistics, Raw Validities and Validities Corrected for Range Restriction for ASVAB Mechanical Composite Standardized Scores and FSG for Mechanical Technical Schools.

AFSC	N	Predictor Mean	Predictor S.D.	Criterion Mean	Criterion S.D.	r	Cor.r[a]	Aj.r[2b]
36131	251	226.48	16.66	88.12	4.25	0.41**	0.63	0.16
41131A	346	225.91	16.81	87.81	4.85	0.37**	0.60	0.14
45234A	803	230.22	17.25	85.41	6.04	0.34**	0.54	0.11
45234B	860	227.95	17.02	85.33	5.66	0.37**	0.58	0.14
45234C	227	229.34	17.53	84.29	5.63	0.39**	0.59	0.15
45430A	920	225.43	20.09	88.60	7.98	0.26**	0.40	0.07
45430B	318	235.70	17.10	89.37	5.38	0.47**	0.68	0.22
45433	311	228.98	17.48	88.64	5.26	0.33**	0.55	0.10
45434	734	230.77	16.43	88.63	9.58	0.11**	0.19	0.01
45730A	109	228.24	17.67	83.24	6.08	0.21*	0.39	0.04
45730B	232	232.61	16.88	83.91	6.24	0.38**	0.61	0.14
45730C	780	227.94	17.23	83.80	6.31	0.42**	0.64	0.17
45730D	120	232.13	19.05	84.21	6.76	0.38**	0.52	0.13
45731	241	230.32	15.25	83.81	6.15	0.44**	0.66	0.19
45732A	544	228.82	17.10	85.08	6.17	0.35**	0.54	0.12
45732B	275	228.79	16.21	84.17	5.86	0.34**	0.56	0.11
45732C	606	228.27	17.65	84.25	6.35	0.37**	0.58	0.14
45830	186	229.80	18.82	88.54	8.55	0.21**	0.37	0.04
45832	870	227.07	17.54	88.63	6.95	0.20**	0.36	0.04
45833	226	218.19	19.22	89.12	7.52	0.12	0.18	0.01
46330	288	233.88	14.35	95.26	2.71	0.22**	0.43	0.05
47230	233	230.74	16.30	89.18	4.34	0.38**	0.62	0.14
47232	504	225.93	17.52	89.47	5.00	0.37**	0.59	0.14
55130	245	223.54	18.37	86.85	4.93	0.36**	0.57	0.12
55131	444	228.83	19.67	89.32	4.72	0.46**	0.63	0.21
55230	313	225.30	15.83	85.78	4.50	0.25**	0.47	0.06
55232	200	220.20	19.15	87.49	4.52	0.36**	0.60	0.13
55235	264	224.39	17.49	84.93	5.03	0.34**	0.57	0.11
56631	240	217.10	18.48	86.07	5.47	0.31**	0.51	0.09

*Correlation significant at $p<.05$
**Correlation significant at $p<.01$
[a]Correlation corrected for range restriction
[b]Coefficient of determination adjusted for population estimate

Table 10.
Descriptive Statistics, Raw Validities and Validities Corrected for Range Restriction for ASVAB Administrative Composite Standardized Scores and FSG for Administrative Schools.

AFSC	N	Predictor Mean	Predictor S.D.	Criterion Mean	Criterion S.D.	r	Cor.r[a]	Aj.r[2b]
27132	137	173.84	8.86	81.40	6.49	0.21*	0.45	0.04
3S031	108	168.78	9.43	90.30	5.43	0.29*	0.54	0.08
49231	246	170.07	8.54	85.65	5.87	0.22**	0.48	0.05
60230	326	169.61	8.84	84.70	5.53	0.01	0.11	0.00
60231	348	169.48	9.13	83.96	5.33	0.08	0.33	0.00
62330	736	165.00	11.92	90.30	5.40	0.10**	0.35	0.01
67231	376	174.11	7.46	85.23	5.34	0.18**	0.40	0.03
67232	441	173.74	7.34	83.90	5.84	0.22**	0.45	005
70230	1988	169.86	10.09	87.96	5.31	0.18**	0.45	0.03
73230	1151	170.34	9.05	82.85	5.75	0.22**	0.50	0.05

*Correlation significant at $p<.05$
**Correlation significant at $p<.01$
[a]Correlation corrected for range restriction
[b]Coefficient of determination adjusted for population estimate

Table 11.
Descriptive Statistics, Raw Validities and Validities Corrected for Range Restriction for ASVAB Administrative Composite Standardized Scores and FSG for Administrative Schools.

AFSC	N	Predictor Mean	Predictor S.D.	Criterion Mean	Criterion S.D.	r	Cor.r[a]	Aj.r[2b]
27132	137	110.18	8.25	81.40	6.49	0.34**	0.48	0.11
3S031	108	109.67	7.96	90.30	5.43	0.40**	0.73	0.15
49231	246	109.31	8.63	85.65	5.87	0.44**	0.63	0.19
60230	326	109.29	7.88	84.70	5.53	0.20**	0.32	0.04
60231	348	109.04	8.07	83.96	5.33	0.34**	0.36	0.11
62330	736	107.89	8.36	90.30	5.40	0.25**	0.28	0.06
67231	376	112.47	7.84	85.23	5.34	0.39**	0.54	0.15
67232	441	111.93	8.28	83.90	5.84	0.36**	0.38	0.13
70230	1988	109.39	8.00	87.96	5.31	0.40**	0.54	0.16
73230	1151	109.93	8.20	82.85	5.75	0.44**	0.45	0.19

**Correlation significant at $p<.01$
[a]Correlation corrected for range restriction
[b]Coefficient of determination adjusted for population estimate

Table 12.
Descriptive Statistics, Raw Validities and Validities Corrected for Range Restriction for ASVAB General Composite Standardized Scores and FSG by Technical School.

AFSC	N	Predictor Mean	S.D.	Criterion Mean	S.D.	r	Cor.r[a]	Aj.r[2b]
11430	296	113.29	5.72	86.98	5.98	0.48**	0.69	0.23
11630	107	110.98	6.96	87.03	5.67	0.51**	0.72	0.25
11730	123	113.95	6.77	90.14	4.76	0.29**	0.38	0.07
12230	541	108.71	7.62	88.16	6.23	0.25**	0.43	0.06
20130	208	115.23	6.84	87.79	7.57	0.17*	0.33	0.03
20230	221	117.00	5.37	93.00	7.02	0.11	0.27	0.01
20530	102	118.16	4.08	88.07	4.76	0.32**	0.68	0.10
20630	128	116.91	5.05	87.66	4.01	0.44**	0.66	0.19
20832A	165	120.98	4.16	87.22	3.30	0.27**	0.64	0.07
20833A	204	121.29	4.31	91.67	2.66	0.27**	0.65	0.07
20834G	104	121.95	3.66	89.75	3.23	0.19	0.53	0.03
27230	814	114.77	6.46	83.88	5.61	0.37**	0.56	0.14
27430	244	109.16	6.58	88.43	5.06	0.42**	0.65	0.17
27530	155	112.43	7.61	86.77	4.62	0.49**	0.67	0.24
27630B	103	112.63	6.75	88.90	6.84	0.52**	0.68	0.26
27630C	221	113.86	6.32	87.60	5.72	0.36**	0.51	0.12
39130	107	114.26	6.65	86.00	5.71	0.45**	0.64	0.20
39230	192	110.83	7.52	84.68	8.98	0.04	0.12	0.00
3C031	103	113.50	7.48	93.03	4.36	0.35**	0.45	0.11
45831	168	110.47	7.48	88.00	5.52	0.34**	0.50	0.11
49131	1865	119.10	5.15	87.61	5.04	0.33**	0.64	0.11
49132	404	123.17	3.45	88.05	5.06	0.30**	0.73	0.09
55330	231	116.26	7.14	83.40	4.79	0.49**	0.65	0.23
57130	1799	109.66	7.16	91.41	9.50	0.07**	0.13	0.01
62330	626	105.91	6.61	89.37	5.79	0.25**	0.51	0.06
65130	269	118.34	4.37	87.32	5.16	0.33**	0.64	0.11
75330	145	109.79	7.26	87.09	5.50	0.42**	0.53	0.17
81130	4874	108.15	7.19	80.18	6.59	0.37**	0.58	0.14
81132	2464	109.05	7.43	80.89	6.21	0.38**	0.57	0.14
90130	307	110.12	7.21	86.21	4.62	0.39**	0.58	0.15
90232	181	109.61	7.84	86.35	5.62	0.51**	0.66	0.26
90330	231	110.97	7.01	85.41	5.25	0.38**	0.56	0.14
90530	323	110.85	7.64	87.82	5.60	0.46**	0.65	0.21
90630	811	109.44	6.94	86.01	5.76	0.33**	0.56	0.11
90730	202	110.73	7.05	85.10	4.41	0.49**	0.63	0.23
90830	203	111.01	6.76	87.71	3.70	0.37**	0.61	0.14
91530	278	110.00	6.75	83.76	5.75	0.40**	0.63	0.16
92430	394	114.86	6.48	86.12	4.59	0.35**	0.55	0.12
92630	225	108.23	6.36	90.03	5.30	0.31**	0.54	0.09
98130	613	109.38	6.49	86.98	5.11	0.34**	0.57	0.11
98230	104	116.21	4.63	86.19	5.17	0.38**	0.71	0.14

*Correlation significant at $p<.05$
**Correlation significant at $p<.01$
[a]Correlation corrected for range restriction
[b]Coefficient of determination adjusted for population estimate

Table 13.
Descriptive Statistics, Raw Validities and Validities Corrected for Range Restriction for ASVAB Electronic Composite Standardized Scores and FSG for Electronic Schools.

		Predictor		Criterion				
AFSC	N	Mean	S.D.	Mean	S.D.	r	Cor.r[a]	Aj.r[2b]
27730	167	235.98	13.09	85.52	5.85	0.40**	0.71	0.16
30333	254	239.30	11.10	90.56	3.85	0.42**	0.71	0.18
30430	411	235.66	12.61	89.14	4.58	0.42**	0.69	0.18
30434	552	234.34	12.31	89.59	4.74	0.52**	0.78	0.27
30436	103	238.86	13.70	89.07	4.64	0.62**	0.81	0.38
30534E	184	239.22	13.36	89.45	4.16	0.50**	0.71	0.24
30636	191	235.41	13.19	89.22	4.46	0.48**	0.71	0.23
32430	350	238.17	12.82	87.92	4.95	0.50**	0.75	0.25
36231	134	230.68	12.58	86.82	5.30	0.50**	0.76	0.25
36234	178	231.05	13.61	87.24	5.21	0.50**	0.73	0.24
41130A	229	236.53	12.07	87.95	4.56	0.49**	0.77	0.24
41132A	238	231.26	14.01	88.43	4.96	0.45**	0.66	0.20
45134A	125	239.61	12.07	93.39	3.38	0.33**	0.52	0.10
45134B	129	237.23	12.62	92.91	3.30	0.48**	0.74	0.22
45135	104	238.83	12.91	93.21	3.34	0.25**	0.47	0.05
45231A	140	237.14	14.36	92.71	3.66	0.35**	0.57	0.12
45232A	145	233.13	12.56	93.16	3.26	0.45**	0.72	0.20
45232B	115	236.01	12.12	91.57	3.46	0.27**	0.52	0.07
45530A	192	240.73	10.65	91.42	4.12	0.35**	0.69	0.11
45531A	326	236.98	13.19	89.00	4.49	0.43**	0.67	0.18
45531B	225	236.98	12.67	89.36	4.26	0.33**	0.59	0.11
45532A	362	236.03	12.96	89.73	4.44	0.52**	0.77	0.27
45532B	246	235.85	12.83	89.84	4.06	0.52**	0.78	0.27
45631A	281	237.95	12.90	90.52	4.23	0.52**	0.77	0.27
45631B	133	235.90	13.47	89.08	4.44	0.54**	0.79	0.28
46630	220	237.25	12.17	88.66	4.48	0.47**	0.76	0.22
49330	419	236.27	12.52	87.25	4.53	0.40**	0.69	0.16
54230	111	226.60	16.30	84.72	5.66	0.43**	0.61	0.18
54231	152	234.01	12.77	85.54	4.98	0.51**	0.75	0.25
91830	111	238.10	12.71	86.22	4.07	0.25**	0.51	0.06

**Correlation significant at $p<.01$
[a]Correlation corrected for range restriction
[b]Coefficient of determination adjusted for population estimate

THE SYSTEMATIC EVALUATION OF ARTERIAL BLOOD PRESSURE REGULATION THROUGH THE ASSESSMENT OF BARORECEPTOR SENSITIVITY AND RESPONSIVENESS TO LOWER BODY NEGATIVE PRESSURE, CAROTID NECK SUCTION, AND INTRAVENOUS INFUSION OF ADRENERGIC AGENTS

Louis Anthony Hudspeth
Graduate Student
Department of Kinesiology and Health Education

The University of Texas at Austin
Bellmont Hall Rm #222
Austin, Texas 78712

Final Report for:
Graduate Student research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Armstrong Laboratories

THE SYSTEMATIC EVALUATION OF ARTERIAL BLOOD PRESSURE REGULATION THROUGH THE ASSESSMENT OF BARORECEPTOR SENSITIVITY AND RESPONSIVENESS TO LOWER BODY NEGATIVE PRESSURE, CAROTID NECK SUCTION, AND INTRAVENOUS INFUSION OF ADRENERGIC AGENTS

Louis Anthony Hudspeth
Graduate Student
Department of Kinesiology and Health Education
The University of Texas at Austin

Abstract

High blood pressure (hypertension) is a condition affecting more than 50 million Americans. Only 10% of the cases are the result of a known etiology. The remaining cases are classified as essential hypertension, meaning that the cause of the condition can not be identified. Prior to the development of successful interventions for hypertension, it is essential that we understand the mechanisms responsible for the regulation of arterial blood pressure.

Blood pressure regulation is a function of the autonomic nervous system (ANS). The integrated baroreflexes (cardiopulmonary, carotid - cardiac, and aortic) play a major regulatory role during orthostatic challenge. The development of experimental protocols which systematically separate these regulatory centers and evaluate their efficacy as isolated entities is essential to furthering our understanding in this area.

The cardiopulmonary baroreflex increases vascular resistance in response to reductions in central venous pressure (CVP), resulting in the maintenance of systemic blood pressure during orthostatic challenge. An LBNP protocol is used to reduced central venous volume thereby causing a reduction of CVP. Reflex sensitivity and responsiveness are evaluated through changes in forearm vascular resistance resulting from the reduction of CVP. The carotid - cardiac baroreflex reduces heart rate thereby maintaining systemic pressure in response to increases in carotid distending pressure. A specialized neck chamber device covers the anterior two thirds of the neck and provides negative pressure to the carotid - cardiac baroreceptors. The stimulus response relationship of the carotid -

cardiac baroreceptors can be obtained by plotting R-R intervals against their respective neck chamber pressure. Isolation of the aortic baroreflex involves the use of LBNP, carotid neck suction, and phenylephrine (PE) infusion. The infusion of PE provides systemic loading of the baroreceptors and thereby serves as an index to integrated baroreflex - cardiac responsiveness. The application of LBNP and neck chamber pressure attenuates the loaded condition of the cardiopulmonary baroreceptors and the carotid - cardiac baroreceptors respectively. Thus, isolating the effects of loading the aortic baroreceptors.

Previous investigators have used these techniques to successfully examine baroreceptors sensitivity and responsiveness to various perturbations. Furthermore, their data suggests that these methodologies are proven and highly reproducible. It is evident from this discussion that additional research is needed to further our understanding of the integrated control of systemic blood pressure. Future research will be instrumental in designing intervention programs for the treatment of hypertension.

Introduction

Hypertension (high blood pressure) is one of the most common health problems in the United States, affecting 18% of the adult white population and 35% of the adult African American population in the United States (Roberts and Rowland 1981). It is one of the leading causes of death in the African American population and a major cause of morbidity and mortality in the Caucasian population (Association 1993). It is estimated that 50 million Americans aged 6 years and older are hypertensive (Association 1995). The Joint National Committee on the Detection, Evaluation, and Treatment of High Blood Pressure defines hypertension as a resting systolic blood pressure [SBP] of $\geq$ 140 mmHg and/or a diastolic blood pressure [DBP] of $\geq$ 90 mmHg (Joint National Committee on the Detection 1993). Of the millions who suffer from this condition, less than 10% can be directly attributed to pathological conditions (Kilcoyne 1980). The remaining cases are classified as essential hypertension meaning that no specific causes can be identified. Prior to the development of successful interventions for hypertension, it is essential that we understand the mechanisms responsible for the regulation of arterial blood pressure. Furthermore, it is imperative to understand the etiology supporting the maintainance of blood pressures in an elevated state.

Blood pressure regulation is a function of the autonomic nervous system (ANS). The centers of the ANS function to regulate heart rate, stroke volume, and total peripheral resistance (TPR) thereby maintaining arterial pressure at rest and during an orthostatic challenge. Since blood pressure is the product of cardiac output (Q) and TPR, evaluation of the systems by which these factors (Q and TPR) are controlled is essential to understanding the regulation of arterial blood pressure. The development of experimental protocols which systematically separate regulatory centers and evaluate their efficacy as isolated entities is essential to furthering our understanding in this area. Thus, the focus of this discussion is to examine the experimental methods used in the evaluation of these

regulatory centers and characterize their ability to define systemic blood pressure regulation.

Cardiopulmonary Baroreflex

The cardiopulmonary baroreflex alters peripheral resistance in response to changes in central venous pressure (CVP). In the presence of an orthostatic challenge central venous volume is reduced. This reduction in volume is accompanied by the reduction of CVP and subsequently stimulates the cardiopulmonary baroreceptors to increase peripheral resistance in an attempt to return CVP to baseline levels. Sensitivity of the cardiopulmonary baroreflex may be assessed through simultaneous measurement of forearm vascular resistance (FVR) and CVP. FVR is calculated by dividing mean arterial pressure (MAP) by forearm blood flow (FBF) The responsiveness of the cardiopulmonary baroreceptors is defined as the slope of the relationship between changes in FVR and CVP. The greater the increase in FVR for a given change in CVP, the greater the reflex sensitivity.

The stimulus response relationship of the cardiopulmonary baroreflex control of FVR is determined by the methods previously described in detail by Gauer and Seiker (1956) and Mack et al. (1987). A 20 gauge tephlon catheter (Angioset) is inserted into a large antecubital vein of the right arm. Subjects are placed in the lower body negative pressure (LBNP) chamber in the right lateral decubitus position. The right arm was allowed to extend through a door in the table supporting the LBNP chamber. The catheter is then connected to a fluid filled pressure transducer (Baxter Uniflow ™) through which CVP can be monitored throughout the protocol (Gauer and Seiker 1956). While in this position, the valves in the veins of the right arm become incompetent and an unimpeded column of blood results. Experimental evidence (Gauer and Seiker 1956, Gauer, Henry et al. 1956) has demonstrated that under this condition the pressure in the veins of the suspended arm are equivalent to CVP when the pressure transducer is at heart level.

The left arm was then instrumented for venous occlusion plethysmography. A high pressure occlusion cuff is placed at the wrist and inflated to 250 mmHg thus occluding all blood flow to the hand. A low pressure cuff (40 mmHg) is placed on the upper portion of the left arm and inflated for 10 seconds followed by 10 seconds of deflation. The low pressure cuff is sufficient to occlude venous flow upon inflation, yet arterial flow is maintained. Therefore, as the low pressure cuff is inflated to 40 mmHg, blood flow out of the arm is occluded, resulting in an increase in the diameter of the forearm. Percent change in forearm diameter is measured and recorded by a Whitney mercury-in-Selastic strain gauge, which is placed at the point of the largest diameter of the forearm (Whitney 1953). Several measurements of forearm blood flow should be obtained at each level of LBNP. The average of these measurements can then provide one value representing the mean forearm blood flow for each stage.

A stepwise reduction in LBNP, causing a footward fluid shift, alters central venous volume. The actions of the cardiopulmonary baroreflex is to counter this reduction in volume by increasing resistance. Thus the reflex defends against a drop in pressure despite reductions in functional volume. Subjects undergo four two minute stages of LBNP at -5, -10, -15, and -20 mmHg. During each stage of LBNP, CVP is recorded at 20 second intervals and FBF is estimated through venous occlusion plethysmography. The change in forearm diameter as a result of the occlusion of blood flow is used to calculated FVR. The slope of the change in FVR in response to changes in CVP due to LBNP are used to define the responsiveness of the cardiopulmonary baroreflex.

Cartotid-Cardiac Baroreflex

Cardiopulmonary baroreceptors are not the sole mechanism by which systemic arterial pressure is regulated. Carotid-cardiac baroreceptors also play a major role in controlling blood pressure. The reduction in heart rate, associated with stimulation of the carotid-cardiac baroreceptors, results in a reduction in blood pressure by reducing cardiac

output. Alterations in heart rate are evident when the carotid baroreceptors are stimulated through the use of a specialized neck chamber device. This device, previously described in detail (Sprenkle, Eckberg et al. 1986), provides a graded series of positive and negative pressure to the carotid baroreceptors.

The neck chamber is designed to cover the anterior three - fourths of the neck. A silicon rubber molding is affixed to the device to provide a proper seal for both positive and negative pressures. The inner latex diaphragm contains a small central perforation. When properly positioned, the opening allows the diaphragm to adhere to the neck and provide a negative distending pressure to the carotid baroreceptors. The device is secured to the subject with velcro straps extending from the base of the neck to either side of the neck chamber device.

The pressure system for the neck device is a computer controlled nickel bellows. Although the bellows system is limited to finite volumes, when properly fitted, the neck chamber provides an excellent seal. Thus the required volume to complete the protocol is minimal. The position of the bellows is established through the computer controller and the electrocardiogram. The result of this integration is stair-stepped neck chamber pressure concurrent with the signal of the electrocardiogram.

The stimulus profile consists of increasing the pressure in the neck chamber to 40 mmHg for five consecutive beats. Following this increase in pressure, a sequential stepwise 15 mmHg pressure reduction is triggered by R waves of the electrocardiogram until -65 mmHg is applied to the neck chamber. The result of this procedure is a graded reduction in neck pressure that is superimposed on carotid arterial pulses. To eliminate the influence on alterations in vagal outflow as a result of respiration, the neck pressure sequence is applied only during held expiration (Eckberg 1983).

The neck chamber device allows for the direct stimulation of the carotid baroreceptors and assessment of the carotid cardiac baroreflex. This stimulation sequence provides negative distending pressure to the baroreceptors, mimicking an elevation in

systemic pressure. In response to this stimulation the carotid cardiac baroreflex functions to return blood pressure to baseline by reducing heart rate and subsequently cardiac output. This reductions in heart rate is accompanied by a obligatory increase in R-R interval; thus providing a measurable parameter for the assessment of reflex sensitivity. However, actual systemic pressure remains constant throughout the procedure. Therefore, reductions in heart rate resulting from the neck pressure sequence are due to stimulation of the carotid cardiac baroreceptors and independent of other feedback regulatory loops controlling arterial pressure.

Test sessions consist of five successful trains of the aforementioned neck chamber protocol. Each sequence lasts ~15 seconds and individual trials should be immediately discarded if the neck device fails to seal properly or if the subject breaths during the procedure. Assuming the complete transfer of pressure from the neck chamber to the carotid arteries, carotid distending pressure is calculated as systolic blood pressure minus neck chamber pressure during each heart beat. The stimulus response relationship of the carotid cardiac baroreceptors can be obtained by plotting the R-R intervals at each pressure step against their respective carotid distending pressure. Available parameters for the assessment of carotid cardiac baroreflex function include: 1. the range of the R-R interval responses, 2. the maximum and the minimum R-R interval responses, 3. carotid distending pressure at minimum and maximum R-R intervals, and 4. the maximum slope, providing an index of reflex sensitivity (Convertino, Adams et al. 1991).

Aortic Baroreflex

Isolation of the aortic baroreflex involves the use of LBNP, carotid neck pressure, and the infusion of phenylephrine (PE). A 20 gauge tephlon catheter is inserted in a antecubital vein of the left arm for the infusion of PE. The subject is then placed in the right lateral decubitus position as described during the cardiopulmonary baroreflex procedure. Following the instrumentation procedure, subjects are allowand to stabilize for 12 minutes, after which three minute of resting data (HR, MAP, and CVP) are obtained.

The protocol is initiated with a steady state infusion of PE. The infusion rate is increased every two to three minutes until MAP has been elevated 15 mmHg above baseline. Thereafter, the rate of PE infusion remains constant throughout the aortic baroreflex procedure. Following three minutes of data collection at the new steady state level (MAP basleine + 15 mmHg), LBNP is applied to attenuate PE induced loading of the cardiopulmonary baroreceptors. The LBNP procedure (ranging form -5 to -20 mmHg) is designed to reduce central venous volume and return CVP to baseline, thus unloading the cardiopulmonary baroreceptors.

Using the neck chamber device previously described, pressure is applied to the carotid baroreceptors. The appropriate pressure, calculated as 1.4 times the change in mean arterial pressure (MAP) due to the infusion of PE, is chosen to insure complete transmission of pressure across the neck (Ludbrook, Mancia et al. 1977). The purpose of this application is to return the mean carotid sinus transmural pressure to pre-PE infusion values. One minute of data is recorded following pressurization of the neck chamber.

Baroreceptor sensitivity is expressed as the ratio of change in HR and MAP (ΔHR / ΔMAP) as a result of experimental intervention. The infusion of PE provides systemic loading of the baroreceptors and thereby serves as an index to integrated baroreflex - cardiac responsiveness. The application of LBNP attenuates the loaded condition of the cardiopulmonary baroreceptors by returning estimated CVP to pre-PE infusion levels. However, the carotid and cardiopulmonary baroreceptors remain in the loaded condition after the LBNP procedure. Application of the calculated neck pressure returns the carotid baroreceptor loading to pre-PE infusion levels thereby isolating the effects of loading the aortic baroreceptors.

Conclusion

Systemic blood pressure remains relatively constant despite changes in posture or orthostatic challenge. This level of regulation is accomplished through a complexed series of reflexes controlled by the autonomic nervous system. To better study the mechanisms by whch blood pressure is regulated, it is neccessary to separate these reflex centers and study them independently. Venous occlusion plethysmography, lower body negative, carotid neck suction and phenylephrine infusion are several methods by which this goal can be attained.

The cardiopulmonary baroreflex regulates blood pressure by altering vascular resistance. Reflex responsiveness and sensitivity can be assessed by monitoring changes in FVR during an LBNP protocol. LBNP causes a footward fluid shift thereby reducing central venous volume and stimulating the cardiopulmonary baroreflex. Vascular resistance, as calculated from venous occlusion plethysmography, increases to defend against reductions in systemic pressure.

When stimulated, the carotid - cardiac baroreflex regulates blood pressure by controlling heart rate. A specialized neck chamber device, providing negative pressure to the anterior two - thirds of the neck, stimulates the carotid - cardiac baroreflex. The neck chamber provides negative pressure to the carotid arteries, stretching the baroreceptors and simulating an increase in systemic pressure. Pressure steps of the neck chamber are triggered by the R waves of the electrocardiogram. Therefore, neck chamber pressure changes are superimposed on the electrocardiogram. The stimulus - response relationship of the reflex can then be determined by examining changes in R-R intervals in relation to the corresponding change in carotid distending pressure. Additionally, reflex sensitivity can be assessed from the maximum slope of R-R interval response curve.

Aortic baroreflex control of cardiac response is assessed by globally loading all baroreceptors. This loaded condition is then followed by the systematic return of cardiopulmonary and carotid conditions to preload conditions. Thus, isolating aortic

control of cardiac response. Sensitivity of the aortic baroreflex is expressed as the ratio of change in HR and MAP (ΔHR / ΔMAP) between baseline and the experimental condition.

Previous investigators have used the techniques described in this paper to successfully examine baroreceptors sensitivity and responsiveness to various pertubations (Engelke, Doerr et al. 1995, Fritsch, Charles et al. 1992, and Convertino, Doerr et al. 1990). It is evident from this discussion that these methodologies are proven and reproducible. Additional research is needed to further our understanding of the integrated control of systemic blood pressure.

References

Association, A. H. (1993). "1993 Heart and Stroke Facts Statistics." American Heart Association :

Association, A. H. (1995). "Heart and Stroke Facts 1995: Statistical Supplement." American Heart Association :

Convertino, V. A., W. C. Adams, et al. (1991). "Impairment of carotid - cardiac vagal baroreflex in wheelchair - dependent quadriplegic." American Journal of Physiology : R576-R580.

Convertino, V. A., D. F. Doerr, et al. (1990). "Head - down bed rest impairs vagal baroreflex responses and provokes orthostatic hypotension." Journal of Applied Physiology **68**: 1458-1464.

Eckberg, D. L. (1983). "Human sinus arrhythmia as an index of vagal cardiac outflow." Journal of Applied Physiology **54**: 961-966.

Engelke, K. A., D. F. Doerr, et al. (1995). "A single bout of exhaustive exercise affects integrated baroreflex function after 16 days of head down tilt." American Journal of Physiology : R614-R620.

Fritsch, J. M., J. B. Charles, et al. (1992). "Short duration spaceflight impairs human carotid baroreceptor - cardiac reflex responses." Journal of Applied Physiology **73**: 664-671.

Gauer, O. H., J. P. Henry, et al. (1956). "Changes in central venous pressure after moderate hemorrhage and transfusion in man." Circulation Research **4**: 79-84.

Gauer, O. H. and H. D. Seiker (1956). "The continuous recording of central venous pressure from an arm vein." Circulation Research **4**: 74-78.

Joint National Committee on the Detection, E., and Treatment of High Blood Pressure (1993). "The fifth the report of the joint national committee on the detection, evaluation, and treatment of high blood pressure." Archives of Internal Medicine **153**: 154-183.

Kilcoyne, M. M. (1980). "The developing phase of primary hypertension II." Modern Concepts of Cardiovascular Disease **49**: 25-30.

Ludbrook, J., G. Mancia, et al. (1977). "Factors influencing the carotid baroreceptor response to pressure changes in a neck chamber." Clinical Science and Molecular Medicine **51**: 347-349.

Roberts, J. and M. Rowland (1981). "Hypertension in adults 25-74 years of age: United States 1971-1975." Vital Health Statistics **221**: 1671.

Sprenkle, J. M., D. L. Eckberg, et al. (1986). "Device for rapid quantification of human carotid baroreceptor - cardiac reflex responses." Journal of Applied Physiology **60**: 727-732.

Whitney, R. J. (1953). "The measurement of volume changes in human limbs." Journal of Physiology London **121**: 1-27.

No report was not available at the time of publication.

CALCIUM CARBONATE SCALE AMELIORATION USING MAGNETIC WATER TREATMENT DEVICES

Kevin M. Lambert
Graduate Student
Department of Civil and Environmental Engineering

Brigham Young University
368 Clyde Building
Provo, Utah 84602

Final Report for:
Graduate Student Research Program
Armstrong Laboratory, Environics Directorate

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington DC
and
Armstrong Laboratory, Environics Directorate
Tyndall Air Force Base, Florida
and
Air Force Civil Engineer Support Agency,
Tyndall Air Force Base, Florida

September, 1996

CALCIUM CARBONATE SCALE AMELIORATION USING MAGNETIC WATER TREATMENT DEVICES

Kevin M. Lambert
Graduate Student
Department of Civil and Environmental Engineering
Brigham Young University

Abstract

The magnetic treatment of water to inhibit scale formation and to remove existing scale deposits continues to engender controversy among water treatment professionals and researchers. However, if magnetic water treatment for scale inhibition were proven, even in limited applications, there would be significant economic and environmental benefits to its use in industry. Even if magnetic treatment for scale amelioration is successfully demonstrated, enough must be understood of its causative mechanisms or its window of applicability so that it may be successfully incorporated into operating industrial systems.

This summer research final report summarizes efforts to find answers to some of the many questions still existing on the topic of magnetic water treatment for scale amelioration. The report begins with an extensive literature review to focus on some of the proposed mechanisms; the problem areas; the best parameters to measure; and methodologies to measure potential changes in calcium carbonate crystals. This review of historical US work, international research, and examples of both successful and unsuccessful applications provides a context for understanding the controversy. Explanations are proposed for the wide diversity of results experienced in both laboratory studies and field trials. An introduction to the large number of variables that affect magnetic water treatment is briefly discussed. A summary of the proposed mechanisms of how magnetic treatment affects scale formation is listed. Recommendations for testing magnetic devices are distilled from successful tests.

The remainder of the report summarizes the system design requirements, planned examination techniques, test plan and design of a test system for examining some of the questions derived from the literature review. This test system will be built and used for parameter testing in the immediate future.

CALCIUM CARBONATE SCALE AMELIORATION USING MAGNETIC WATER TREATMENT DEVICES

Kevin M. Lambert

INTRODUCTION

This report summarizes the research conducted during the summer of 1996 at Environics Directorate on the subject of magnetic water treatment devices used to prevent or reduce calcium carbonate scale formation on industrial piping and heat exchange surfaces. The majority of the report is a brief overview of an extensive literature search. The remainder of the report summarizes the design requirements, planned examination techniques and design of a test system for examining some of the questions derived from the literature review.

Scale prevention and reduction through the use of magnetic water treatment devices has been a controversial topic for many years and remains so today. Many engineers and scientists believe that these devices do not perform as claimed by marketers of these devices, while the proponents of this technology make many diverse (some might say extravagant) claims about their successful applications. It is difficult to know what to believe about these devices by reviewing the published literature due to the many conflicting results. Treatment by electro- or permanent magnets is one of several non-chemical means proposed for treating scaling waters, which if successful, even in limited applications, would provide significant economic and environmental benefits. Much of the removal of scale deposits in industrial heat transfer equipment is accomplished through acid washes. Reduced scaling would reduce the use of these acids. Another benefit of reduced scaling in boilers is lower power plant emissions due to lower fuel consumption corresponding to more efficient heat transfer.

The purpose of the literature review is to give a brief overview of laboratory research and field experience with magnetic water treatment devices (MTDs) both in the United States and in foreign countries over about the last 45 years. The concentration is on research from

the last dozen years. While some of the information is applicable to many forms of scale, the information presented here focuses on calcium carbonate scale specifically. The review is broken down into the following topics: brief history; experimental results; parameters affecting magnetic device testing; unsuccessful and successful magnetic device tests; why there are so many conflicting results; classifications of proposed mechanisms; and recommendations for device testing. This literature overview provides a framework for understanding why there are so many conflicting reports regarding the operation of these devices.

BACKGROUND: LITERATURE OVERVIEW

Brief History

There is very little in the open literature in the U.S. prior to the 1950s regarding the examination of magnetic water devices. In the first half of the 1950s several U.S. engineers and scientists wrote articles attacking the statements of sales literature prevalent at that time, but no attempt was made to test the devices. The second half of the 1950s saw several serious attempts by researchers to test scale-preventing magnetic devices. None of the tests showed any success in preventing scale formation. There is very little U.S. published literature on the subject in the U.S. between 1960 and 1977. I believe the device testing of the late 1950s convinced many that the devices did not work. However, many articles came out in Europe and the former Soviet Union during the 1970s, generally indicating from moderate to considerable success in reducing adherent scale formation and removal of existing scales using magnetic water treatment devices. Field testing, as reported by water-treatment magnetic-device-marketing companies and occasionally by customers, has continued to show successful applications of these devices. Unsuccessful field trials were rarely reported by these sources.

Starting in the late 1970s serious independent research in the U.S. began again to examine the effectiveness of magnetic treatment of water to prevent scaling. Until the mid 1980s essentially all the independent laboratory tests and field trials showed little or no effect on measured parameters due to the use of MTDs. Since the mid 1980s more U.S.(22) and

foreign (7) researchers have found significant, measurable changes in several calcium carbonate crystal parameters. Scale reduction has been verified in some instances (8) and scale removal has been reported occasionally. However, some research has continued to show no measurable changes in water characterization parameters or scaling due to the use of magnetic devices. (12,17) The following list includes the majority of reported successful applications of these devices in reducing adherent scale: boilers, cooling towers, steam generators, air-conditioning condensers, sugar-processing plants, oil field production, and residential hot water heaters. (1)

Experimental Results

The most obvious questions examined in the literature have been whether magnetic water conditioning devices reduce scale formation on pipe or heat exchange surfaces and whether they remove or "soften" existing scale from these surfaces. Many other water and calcium carbonate crystal parameters have been examined as part of the effort to prove or disprove the claimed phenomenon and to understand underlying mechanisms that may explain its functioning. Measuring parameters other than direct formation of scale not only helps in the search for understanding the phenomenon, but in some cases is a quicker and easier means to look for magnetic effects in the aqueous solutions tested. The listing here will provide a quick look at the various parameters examined in the published literature.

Scale surface deposition: Some research and field trials have shown success in reducing scale formation and some have even shown reduction in existing scale deposits.(7)

Corrosion: The use of an MTD has been reported to increase corrosion of steel and (8,11) iron. Other data suggest inhibition of iron or steel corrosion due to the presence of an operating MTD. (1) No consensus has been reached about the effect on iron and steel. Data show increased corrosion for active state titanium but reduced corrosion of aluminum and zinc due to the presence of operating MTDs. (1)

Electrical properties: One report shows that voltage and current changes can be measured in conducting fluids treated with MTDs relative to the same fluids operating without MTDs.(4)

Crystal phases: This has been a significant area of research on the question of anti-scale magnetic treatment (AMT) of water. Many researchers in different countries have reported measurable changes in the calcium carbonate crystal phase.(1,21) Calcium carbonate is frequently found in two polymorphic forms, which are identical in chemical composition, but differ in crystal shape and density. These two crystalline phases are calcite and aragonite. A third crystalline phase, vaterite, is infrequently found. The changes most commonly reported in the literature for precipitated calcium carbonate crystals are noted below.

Crystals precipitated from aqueous solutions without AMT are composed principally of calcite (8)(70 - 80% is the most commonly reported range),(6,7) the remainder being aragonite. After the solutions flow through MTDs and precipitated crystals are examined, they are found to be primarily aragonite (14)(70 - 80% has been reported by several publications) with the balance composed of calcite. Adherent scale removed from pipe and heat exchanger surfaces has generally been determined to be composed mostly of the calcite phase. Precipitated crystals removed from the bulk fluid (by filtration or settling in quiescent zones) generally have been shown to be mostly aragonite. With different crystalline shapes, densities, and ions that can substitute into the respective crystal lattices for calcite and aragonite, there are some significant differences between these two phases. Some researchers believe that this noticeable effect is tied to the scale reduction phenomenon.

Other crystal factors: Other changes in the precipitated crystals that have been noted include size, number and crystal shapes. While published results have shown increases and decreases in both crystal size and number, it appears that the majority of the reports favor an increase in crystal size (6) accompanied by a decrease in crystal numbers (14) due to the effect of AMT. Many changes in crystal shape after AMT have been reported. (16)

pH: (4,11) While many researchers have reported little or no change in pH due to AMT (11), one did measure a pH change of 0.5 by controlling different test parameters than did other researchers.

Zeta potential: Few researchers have measured zeta potential, but this parameter indicates a potentially powerful argument for changes due to AMT. Twenty-five percent is the maximum reduction in Zeta potential measured for a solution treated by an MTD. Reduced potential allows charged particles closer proximity, facilitating coagulation of colloid particles. (4,11)

Impurities: Some researchers have argued that reduced scaling due to the use of MTDs derives solely from the presence of certain known scale -reducing ions, especially iron. These researchers proposed that corrosion of the MTD itself or of the adjacent pipes supplied the small concentrations of iron necessary to suppress scale formation. Other researchers argue that iron and various colloids are necessary for the successful application of AMT. They showed that the use of AMT with small concentrations of iron and colloids reduced scale formation significantly more than without AMT. Thirty-four chemicals were tested in the mid 1980s in one study (23) alone for their effect on calcium carbonate crystal growth kinetics. Some impurities are used industrially as scale suppressants. (11)

Solubilization rate: One study showed the solubilization rate of calcium carbonate to increase as much as 43% due to the use of MTDs. (1)

Conductivity and dissolved solids: Both of these parameters have been measured at less than a 10 % reduction due to AMT. Some tests have shown no change to these parameters. (1)

Suspended solids and infrared absorbance: Some tests showed no change to these two (3) parameters. Other tests have shown a significant (25-30 %) change in value due to magnetic

treatment. Some later researchers proposed that the significant changes measured were the result of the presence of impurities not noted by the those observing these larger changes.

Physical water parameters: No significant changes have been reliably measured in many physical water characteristics such as density, viscosity, boiling and freezing points, visible light transmission and reflection. (19)

Memory effect: This is a very important and characteristic feature that always shows up when magnetic treatment has been reported to produce significant, measurable changes. Whatever characteristic or parameter produces a measurable change is shown to persist for several hours up to about a week after magnetic treatment is terminated. (2,14,21) This is both an important practical effect for successful AMT and tied to understanding the underlying mechanism.

Parameters Affecting Magnetic Device Testing

A large number of factors have been reported by one or more authors to have a significant effect on the testing of MTDs. They are briefly introduced here to indicate the types of factors that must be controlled or measured for successful testing of MTDs. Successful results as used here solely indicate that AMT was able to demonstrate a significant, measurable change in the parameters examined. It does not necessarily mean that scale deposition was noticeably reduced, as this parameter was not always measured.

Calcium carbonate saturation level: This is the most-commonly accepted requirement (19,20) for an MTD device to show successful results. The solution must be supersaturated with respect to calcium carbonate at the time and point of application of the magnetic device. The supersaturated condition is determined using the Langelier Saturation and Ryznar Indices.

Magnetic field strength or intensity: Several reports show that increasing magnetic field strength increases whatever (2,14,19) effect is being measured up to a cutoff point. This point of no additional effect occurred about 0.3 to 0.5 tesla (T) (3000 - 5000 gauss (G)).

Magnet design and field orientation: (2) Electromagnets are commonly used in the former Soviet Union but have been infrequently investigated in this country. Promoters of MTDs defend the importance of different arrangements of permanent magnets which include pole arrangement and spacing. Whatever the design, the magnetic force lines should be perpendicular to the flow velocity. This produces the largest Lorentz forces induced by the magnetic field. Lorentz forces are thought by some to be the causative factor underlying the magnetic effect. (1)

Magnet installation: Another possible effect is whether the magnet is installed in-line (the solution flows around the surface of the magnet) or whether it is installed external to any pipes. The in-line style produces flow blockage and turbulence (thought by some to assist the magnetic effect or coagulation process) but is more difficult to install and remove. In-line may also introduce chemical effects (corrosion) which may add or obscure scaling mechanisms.

Wetted surfaces: The piping and heat exchanger construction materials may affect test results if they supply small quantities of impurities that affect scale formation or crystal nucleation or growth kinetics. Different surface finishes also affect crystal nucleation on the solid surfaces. For example: scale does not adhere as readily to the smoother surfaces of PVC pipes. (5)

Time effects: The total exposure time of the fluid to the magnetic field has been shown numerous times to affect the outcome of AMT tests. (14) The exposure time is influenced by fluid velocity, number and length of the devices used and the number of passes recirculated water makes through the magnetic field. Also important is the length of time since magnetic exposure before a solution is examined. This is tied to the memory effect. (16)

Fluid properties: Fluid temperature and pH very significantly affect the solubility of calcium carbonate. (16) Fluid pressure is significant only in highly pressurized systems.

Flow conditions: Flow velocity affects the magnetic exposure time and the magnitude of the Lorentz forces. High velocities can affect crystal nucleation on sidewalls and can produce a scouring effect, limiting the total adherent scale thickness. Several published reports indicate an influence due to fluid turbulence, whether due to the system design or fluid velocity or artificially created by an in-line magnet. The Russians especially have commented on this factor. Some results indicate successful AMT above the laminar range. If more than one phase is present in the flowing solution, crystal nucleation can be impacted. Nucleation is affected by vapor- - liquid interfaces such as vapor bubble surfaces.

Impurities: Many impurities, some at very small concentrations, have a large impact on crystal growth kinetics. Many inorganic and some organic impurities (9,15) have an effect, mostly to inhibit crystal growth rates. Different impurities substitute into the calcite and aragonite crystal structures, affecting both their growth rates and transformations between the two phases.(1,13)

Heat load / specific heat rate: A few researchers have shown the rate of heat transfer supplied by the heat exchange equipment can significantly affect the AMT effect on scaling. (12,19)

Specimen preparation: One of the popular techniques for examining calcium carbonate crystals is X-ray diffraction (XRD). Grinding and storage of the scale specimens can affect the composition of the crystal phase measured (calcite vs. aragonite).

Measurement methodologies: The measurement methodologies used don't change the crystal parameters affected by the use of AMT, but in some cases may change the interpretation of the noted results. Specimen preparation is one example of this phenomenon.

"Unsuccessful" and "Successful" Magnetic Device Tests

Examining specific examples of both "successful" and "unsuccessful" laboratory tests or field trials in the literature can be very instructive in understanding why there are so many conflicting results and conclusions reported. It is very important to look at how the tests were conducted, what parameters were measured, and how the results were interpreted.

Controlled tests were run on both non-magnetic and magnetic water treatment devices in tube heat exchangers between 1975 and 1984. (18) Two electromagnetic devices and two permanent magnetic devices were tested. The published report concluded that none of the magnetic devices significantly reduced scale. This is the same conclusion reached by independent laboratory and field tests reported in 1977 and the late 1950s.

The published data for this research showed that two of the MTDs tested showed scale reductions of 14 - 16%. While this is not a large reduction, it is large enough to be confidently measured, and may in fact show successful treatment given the parameters to be discussed next. Several parameters currently considered important in successful AMT applications were in ranges during this research that would indicate at best a very marginal application for successful scale reduction due to AMT. These include very low levels of iron in the treated water, significant temperature variations, a single-pass system with short magnetic exposure times, and problematic calcium carbonate saturation levels. The published data were used to calculate Langelier Saturation and Ryznar Indices. These indicate that the water was likely not supersaturated with calcium carbonate at the point of exposure to the magnetic field and reached marginal supersaturation levels only in the effluent from the heat-transfer equipment. It may well be that the particular conditions of this testing severely limited the potentially successful application of the MTDs used in this study. The small scale reduction of two of the devices may in fact be all they were able to do given the marginal operating conditions.

Now let's review some "successful" applications. The U.S. Coast Guard had a land-based boiler that experienced 40% area reduction in its piping due to adherent scale. An MTD was installed and after several months of operation there was a 41% fuel savings due to reduced boiler fuel requirements, the pipe scale was cleared out, and the exit water temperature increased by more than 20° F. A large quantity of loose, soft scale was removed from a stagnant point in the system. The Coast Guard also applied MTDs to six boilers on six ships. They measured alkalinity, chlorides and scale before and after chemical conditioning was terminated and magnetic treatment was begun. Begun in 1989 and continuing through at least the end of 1992, the Coast Guard was very satisfied with the results.

As with the previously discussed test results, it is instructive to examine the test controls and reporting. In these published reports there was only a small amount of direct comparison of measured test results with and without AMT. The operating water was poorly characterized and there was little direct control of the experiments so it is difficult to say that the MTDs operated under the same conditions as did the chemical treatment. Also, on the land-based boiler, a special blowdown schedule was instituted. This type of blowdown schedule is known to retard scale formation and is a commonly reported procedure used when magnetic device marketers have a say in the operation of the system for comparison testing. So it is difficult to use these reported results to really give AMT a passing grade for scale prevention, although it looked quite convincing.

Why Are There So Many Conflicting Results?

It becomes evident that many reported results from AMT testing have had very different results reported for the same parameters from tests performed by different researchers. I believe that this confusion is due to several factors. 1) There are so many inter-related variables. Different parameters dominate solution chemistry, and crystal nucleation and growth under different operating conditions. 2) Many of the reported tests or field trials indicate a lack of control of many of the influential factors or poor characterization of the

tested water. Some of the tests measured parameters that in fact do not change even under reported successful AMT applications. 3) There is incomplete understanding of the many variables that influence potentially successful applications of AMT. This misunderstanding generally causes the lack of control or characterization of experiments, although sometimes this is due to lack of the ability to control or measure certain parameters due to a particular system configuration or lack of funds for measurement equipment. Two very recent examples serve to illustrate these issues.

A utility power plant attached an MTD to a pipe that carried 1% of the total system flow to a holding lagoon where the water cooled. After several days this water was added to the rest of the system flow. The plant manager reported that the MTD was completely unsuccessful in reducing scale. But an understanding of current research indicates that there are at least three problems with this application as tested. 1) The magnetic field was applied to water just before it entered a lagoon for cooling. The problem: low temperature at this point may have indicated an undersaturated calcium carbonate solution. 2) The several-day time delay may have negated any potentially successful water conditioning by the MTD due to the memory effect. 3) Treated water does not somehow magically cure the rest of the water it is mixed with. So at most, one would have observed no more than a 1% scale reduction (probably not even noticeable) even if the AMT had been 100% effective.

A large government agency is nearing the end of a two year field test of four magnetic devices. It was verbally reported that none of the devices had shown successful results. In particular it was reported that one system was doing so poorly that filtration had to be added to remove all the precipitated calcium carbonate crystals flowing in the fluid. If accurately reported this actually indicates a successful application of AMT. If calcium carbonate is in the water, it can go only three places: 1) remain dissolved in solution, 2) precipitate out as adherent scale or 3) precipitate out as non-adherent crystals that remain in the bulk fluid. If precipitated crystals that remain free floating in a recirculating system are removed with filtration, then the calcium carbonate concentrations in the bulk fluid can gradually be reduced.

If the bulk-fluid concentration is reduced, this condition can lead to dissolution of existing adherent scale on pipe surfaces.

Classification of Proposed Mechanisms

Many mechanisms have been proposed to explain scale amelioration through the use of AMT. These different mechanisms have been organized into two different kinds of classification systems. One classification system groups the theories as follows: A) Interatomic effects, B) Contamination effects, C) Intermolecular / ionic effects, D) Interfacial effects.(1) The other classification system groups the different theories into three different categories: 1) Physical / structural water changes, 2) Effect of iron impurities, 3)Lorentz force effect on ions and colloids.

Recommendations for Device Testing

This is a summary of the parameters more commonly reported in the literature where measurable changes were observed. a) The magnetic field orientation is perpendicular to the flow direction. b) Several reports have indicated more success with magnetic field strengths of a minimum of 0.3 - 0.5 T (3000 - 5000 G). c) There are multiple indications that small concentrations of iron and other colloids (commonly occurring in natural, hard waters) must be present in the tested water. d) General indications are that a higher flow velocity is better than stagnant conditions. e) A recirculating water application is recommended. f) It is believed that the aqueous solution must be supersaturated in calcium carbonate at the point and time of application of the magnetic field. g) Generally there is more success with a little higher temperature waters (above room temperature, generally above 100° F.

A few operational recommendations are proposed: a) Provide filtration (side stream or bypass filtration may work well) to capture non-adherent precipitated crystals and remove them from the system fluids. b) Characterize the water chemistry and system operating

conditions. c) Maintain the same operating conditions (if possible) with and without magnetic devices installed.

Final Comments on Literature Review and Background

Based on recent research it appears likely that there are limited applications in which some magnetic devices will reduce calcium carbonate scale deposition. It is possible that even with these successful applications that some continued chemical treatment may be required at reduced levels. It has been repeatedly shown that not all magnetic devices reduce scale under many conditions. With substantial recent research data showing some successful potential for magnetic water treatment for scale prevention, continued research work should be pursued. The promise this technology holds out for economic and environmental benefits would justify its use as another tool for water treatment if proven effective.

TEST PROGRAM

The remainder of this report summarizes the effort to design a test system and program to pursue examination of some of the issues discussed in the literature review.

TEST GOALS and MEASURES of SUCCESS

The principal goals to be pursued in a test program are: 1) Demonstrate some successful (if it occurs) application of an externally mounted, permanent magnet that shows repeatable, significant, measurable changes in calcium carbonate crystal parameters. 2) Find some correlation between the measured output and the varied inputs. For the proposed test program, success would be defined by significant differences in crystal size or morphology or in aggregate particle topology between non-treated and treated water samples.

TEST SYSTEM DESIGN REQUIREMENTS

Based on the literature review, consultation on Air Force applications and technical and budget constraints, a set of design requirements was developed for a test system to be built. The design requirements and operational parameters selected are as follows:

A) All wetted surfaces are to be non-metallic to minimize sources of impurities.

B) All wetted materials are to be able to sustain temperatures in excess of 110° F under system pressures for the section they are located within.

C) To minimize potential damage to aggregated particles all components (especially the pump) are to be selected to minimize turbulence, crushing or shearing.

D) System pressures are only to exceed atmospheric to the extent necessary for pumping and to provide sufficient head for the sidestream filtration.

E) Due to the complexity and cost of an automatic pH monitoring and self maintaining system, it was decided that pH will be measured, but not automatically controlled. The system will be allowed to come to equilibrium before testing continues.

F) Distilled water will be filtered and passed through ion exchange media to assure that none of suspected contaminants (that significantly impact solubility) are present. Additions will then be mixed in to give desired alkalinity, hardness, etc.

G) The system was sized to approximate pilot plant size (30 gallons).

H) The pump selected will be capable of providing sufficient head for filtration purposes along with a wide range of volumetric flow rates. It will have no wetted metal parts, and its operation will minimize potential breakup of any aggregated particles.

I) Measurement of system pressures, bulk fluid temperature, pH, alkalinity, hardness and certain ion concentratons is required.

J) A method of heating and mixing the bulk fluid in the reservoir is required.

K) The magnetic device is to be externally mounted to prevent flow disturbances, eliminate a potential source of contaminants and for ease of system changes. A permanent magnet is to be used to eliminate the need for a permanent electrical supply (not significant in lab testing but

can be important in real applications). A magnetic field strength of greater than 0.5 T is required (closer to 1 T is preferred).

L) To eliminate the problems associated with a variable output heat exchanger due to large flow velocity fluctuations, the bulk fluid in the reservoir will be heated. Provide system insulation, if required, to maintain a constant temperature.

M) Measuring the effect of AMT through examining precipitated crystals within the bulk fluid provides the advantage of reducing the time of individual test runs, as opposed to measuring scale. This is accomplished by side-stream filtration down to a particle size of 1-2 μm.

N) Allow removal of magnetic field while maintaining system flow to examine memory effect.

O) The particle sampling methodolgy must remove moisture from the samples to prevent aragonite/calcite phase transformation.

P) Pressure gages to be appropriately located to monitor system operation and filter plugging.

PLANNED EXAMINATION TECHNIQUES

Many water physical and chemical parameters will be measured before and after test runs utilizing fluid samples taken from the reservoir. Pressure, temperature, pH, alkalinity and various hardness values will be monitored during the test runs. Particles will be removed from the bypass filter and from the mini-drain at the bottom of the cone-bottomed reservoir. The particles will be examined by scanning electron microscopy (SEM) for general size and shape information. X-ray diffraction (XRD) will be used to determine relative proportions of calcite and aragonite present. Another technique is being considered for examining particle topology.

TEST SYSTEM DESIGN

Piping, fittings and tubing were selected for a minimum use temperature of 125° F at rated pressure and to be compatible with all chemical species anticipated for use. An air-operated double-diaphragm pump with a surge dampener was selected to provide relatively smooth, constant flow over a large range of flow rates. It also presents non-metallic surfaces

and minimal potential damage to aggregated particles in the flow stream. A dual-layer filtration system was selected to screen for both individual crystal sizes and for aggregated particles without rapid plugging of the finer filter. Smooth-surface filter membranes were selected for ease of removal of the particles from the membranes. An overall system schematic is shown in Figure 1.

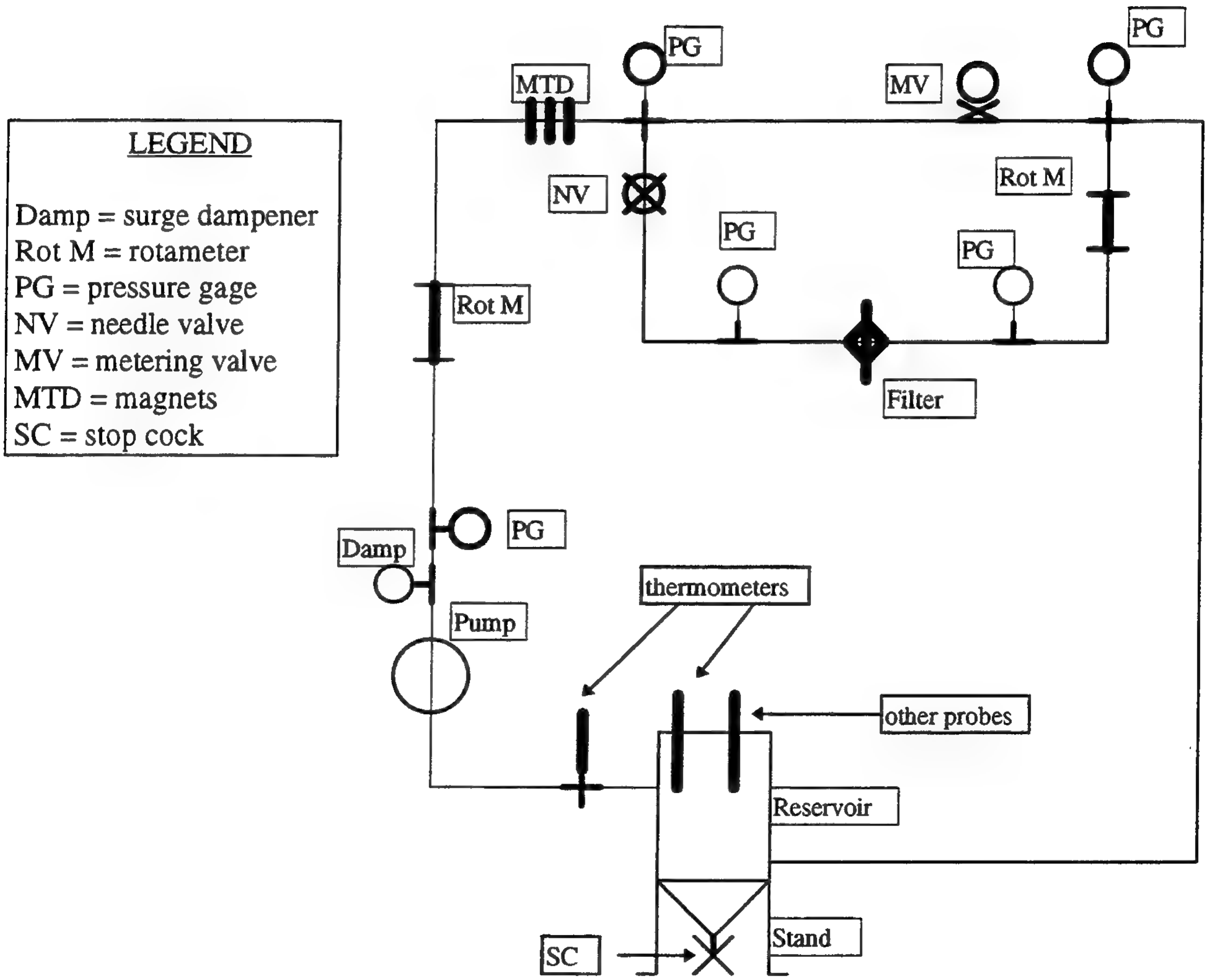


Figure 1 Test System Schematic

REFERENCES

1) Baker,J.S. and S.J. Judd, Magnetic amelioration of scale formation, Water Research, Vol. 30, No. 2, 1996, pp. 247-260.

2) Belova,V., Magnetic treatment of water, Soviet Science Review, May 1972, pp. 150-156.

3) Bernardin, J.D. and S.H. Chan, Magnetic effects on simulated brine properties pertaining to magnetic water treatment, Fouling and Enhancement Interactions, ASME, 1991,HTD-Vol.164, pp. 109-117.

4) Busch, K.W., M.A. Busch, D.H. Parker, R.E. Darling and J.L. McAtee, Jr., Studies of a water treatment device that uses magnetic fields, Corrosion - National Association of Corrosion Engineers, Vol.42, No.4, April 1986, pp. 211-221.

5) Cowan,J.C. and D.J. Weintritt, "Water-Formed Scale Deposits", Gulf Publishing Co.,©1976

6) Deren, E., Le traitement des depots calcaires dans l'eau par le procede CEPI, L'eau, L'industrie, Les nuisances, No. 91, Avril, 1985, pp. 49-52.

7) Donaldson,J.D., Magnetic treatment of fluids, Brewer's Guardian, July 1990, pp.20-24.

8) Duffy,E.A., Investigation of magnetic water treatment devices, Ph.D. dissertation, Clemson University, 1977.

9) Duncan,S.L., Phase and morphological control of calcium carbonate, Master's thesis in Materials, Pennsylvania State University, 1995.

10) Elliasen,R. and R.T.Skrinde, Experimental evaluation of water condtioner performance, Journal American Water Works Association, Vol. 49, No. 9, September 1957.

11) Ellingsen, F. and L. Fjeldsend, A revue of scale formation and scale prevention, with emphasis on magnetic water treatment, Proceedings of the 14th Internationall Water Supply Association: World Congress of the Int'l Water Supply Assoc., 1982, Vol. ss8, pp. 12-19.

12) Hasson, D. and D. Bramson, Effectiveness of magnetic water treatment in suppressing $CaCO_3$ scale deposition, Ind. Eng. Chem. Process Des. Dev., Vol.24, No.3, 1985, pp.588-592.

13) Heffner,J.D., Precipitation of metastable aragonite fromaqueous solution, PhD dissertation in geology, University of Pennsylvania, 1976.

14) Higashitani, K., A.Kage, S.Katamura, K.Imai and S.Hatade, Effects of a magnetic field on the formation of $CaCO_3$ particles, Journal of Colloid and Interface Science, 156,1993, pp90-95

15) Kazmierczak,T.F., Kinetics of precipitation of calcium caronate, PhD dissertation, State University of New York at Buffalo, 1978.

16) Kronenberg,K.J., Experimental evidence for effects of magnetic fields on moving water, IEEE Transactions on Magnetics, Vol. MAG-21, No. 5, September 1985, pp. 2059-2061.

17) Lawrence, D.J., Evaluation of commercial magnetic descalers, CERL Technical Report M-342, US Army Corp of Engineers, May 1984, pp.1-30.

18) Limpert, G.J.C. and J.L. Raber, Tests of nonchemical scale control devices in a once-through system, Materials Performance, October 1985, pp. 40-45.

19) Martynova, O.I. and B.T. Gusev (Moscow Power Institute), The effect of a magnetic field on reduction of the intensity of scale formation as applied to the demineralization of sea water, translation by SCITRAN for Naval Construction Battlion Center, Port Hueneme, CA 93108

20) O'Brien, W.P., On the use of magnetic (and electric and ultrasonic) fields for controlling the deposition of scale in water systems - A review of several(22) papers translated from Russian, Civil Engineering Laboratory of Port Hueneme, CA, October, 1979, pp.1-23.

21) Pandolfo, G.P., Magnetic field and tap water, La Chimica e l'industria, Vol.69, No.11, November 1987, pp. 88-89.

22) Raisen, E., The control of scale and corrosion in water systems using magnetic fields, Paper No.117, Corrosion 84 - N.A.C.E. Conference, April 2-6,1984, New Orleans, LA, pp. 1-20.

23) Meyer, H.J., The influence of impurities on the growth rate of calcite, Journal of Crystal Growth, Vol.66, 1984, pp. 639-646.

A QUANTITATIVE REVIEW OF THE APTITUDE TREATMENT INTERACTION LITERATURE

Robyn M. Maldegen
Doctoral Candidate
Department of Psychology

Texas A&M University,
Department of Psychology
College Station, TX 77843-4235

Final Report for:
Graduate Student Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Armstrong Laboratory

August 1996

A QUANTITATIVE REVIEW OF THE APTITUDE TREATMENT INTERACTION LITERATURE

Robyn M. Maldegen
Doctoral Candidate
Department of Psychology
Texas A&M University

Abstract

Aptitude treatment interactions (ATI) describe the idea that the optimal learning environment for any given person depends on their unique set of aptitudes (Cronbach & Snow, 1981). This study integrates the ATI literture by developing a framework for aptitude treatment interactions and quantitatively reviewing the relevant literature within this framework. Studies were coded based on student level, study design, how the dependent variables were measured, what type of aptitudes were investigated, how aptitudes were measured, the type of instruction manipulated, instructional method, and course content. Frequencies were then calculated for each category. Results indicate that researchers typically examine cognitive aptitudes, conative aptitudes, and affective aptitudes. In addition, comparing a structured to unstructured approach and an elaborative approach to one that provides little additional information were the most common ways to manipulate the treatment. Finally, the review suggests that traditional instructional methods such as, lecture, discussion, practice, and textbook use are still among the most frequently used methods for instruction. Implications from this study will be used to develop a taxonomy for classifying individual characteristic variables for training.

A QUANTITATIVE REVIEW OF THE APTITUDE TREATMENT INTERACTION LITERATURE

Robyn M. Maldegen

Aptitude treatment interactions (ATI) have received a great deal of research attention since the 1960's (Shute, 1992). Research applications have been diverse, ranging from classroom settings (Stensvold & Wilson, 1990) to personnel selection and training (Mumford, Harding Fleishman, & Weeks, 1987) and have been applied to young children (Ysseldyke, 1977), managers and administrators (Gist, Schwoerer, & Rosen, 1989) and the range in between. In addition, both aptitudes and treatments have been conceptualized in a variety of ways. For example, the types of aptitudes studied have ranged from cognitive abilities, such as crystallized intelligence and spatial ability to personality variables, such as self-efficacy and demographic variables, such as sex and race (Cronbach & Snow, 1981). The types of treatments studied have ranged from structured versus unstructured instruction (Snow, 1989) to traditional lecture versus intelligent tutoring (Shute, 1993). One initial literature search to be used in a meta-analysis of ATI's located 250 relevant studies (Kavale & Forness, 1987). It is evident that work on ATI's is diverse and that integration of this research, although quite needed, can be a difficult and often times a confusing undertaking.

The purpose of this study is to provide a means for integrating this immense body of literature by describing one possible taxonomy for aptitude treatment interactions. In addition, I will summarize a portion of the literature quantitatively to demonstrate how the taxonomy can be useful. This paper is intended to be used in the process of

developing a classification system for ATI's in training settings for the Air Force. As such, when summarizing ATI studies, the emphasis will be placed upon issues concerning how aptitudes are defined and measured, what types of treatment levels are investigated, the types of course content studied, and the methods used to convey course information to students.

Aptitude treatment interactions

According to the framework set forth by Cronbach and Snow (1981), aptitude treatment interactions explain the notion that the optimal learning environment for any given individual will depend on their unique set of aptitudes. Conceptually, this is an extremely appealing idea. Of course a student with strong auditory abilities and weak visual abilities will learn better when the lessons are given verbally rather than visually. Right? Although this seems like a reasonable theory, it has been the source of great debate. In searching for potential moderators in a meta-analysis of training effectiveness, Bennett (1996) examined the hypothesis that specific training characteristics or ATI's moderate training effectiveness. The results did not support this hypothesis. Another meta-analysis investigating the efficacy of modal training came to the same conclusion (Kavale & Forness, 1987). The other side of this debate, however, is that so much research is being conducted that researchers are able to fill volumes of books with evidence supporting aptitude treatment interactions (see Ackerman, Sternberg, & Glaser, 1989; Cronbach & Snow, 1981; Regian & Shute, 1992; Sternberg & Wagner, 1994).

In general, the explanations offered to account for inconsistencies in the results of ATI studies have been methodological in nature. For example, in Bennett's (1996) meta-analysis, he notes that most of the ATI studies conducted before 1977 had a sample size

of fewer than 40 subjects. It was also noted in the meta-analysis that random assignment to conditions was typically not conducted. When Cronbach and Snow (1981) conducted their review of the literature it was noted that studies of ATI's tend to lack methodological rigor. The authors provided ample discussion of how to conduct more rigorous investigations of ATI, however, that does not change that it is very difficult to control for all of the factors in a classroom setting that could potentially influence the results (e.g., teacher personality, classroom dynamics).

In order to understand the optimal learning environment for a given set of aptitudes it may be helpful to put aptitudes and interactions into a framework. Shute, Lajoie, and Gluck (in press) discuss three categories of aptitudes: cognitive, conative, and affective. Cognitive factors are defined as "mental processes and structures associated with knowledge and skill acquisition, such as working-memory capacity and general knowledge" (p. 12). Conative factors can be thought of as more stable traits, such as motivation or competitiveness. And closely related, affective factors are less stable moods and personality traits such as, happiness, fatigue, conscientiousness, or neuroticism (Shute et. al., in press). The authors also point out that these factors are ordered in terms of stability of the aptitude with cognitive factors being the most stable in relation to learning outcomes and affective factors being the least stable.

Due to the range of potential cognitive factors, it may also be useful to draw further distinctions between the types of potential cognitive factors. Snow (1994) draws the reader's attention to a study by Carroll (1993), in which a factor analysis of cognitive abilities found that a set of first-order factors can be collapsed into a set of second-order factors including, fluid reasoning, crystallized language, visual perception, auditory

perception, memory, speed, and idea production. Finally, these second-order factors make up the third-order factor, general intelligence.

Snow (1994) also provides a framework for considering treatment levels. Examining multi-dimensional treatment levels is a new concept to most ATI theorists who have only considered one or two levels of a treatment as the independent variable in past research. It is suggested that some distinction be drawn between tasks, treatments, and contexts. Snow (1994) draws from Doyle (1983) who defined four categories of academic tasks: memory tasks, procedural-routine tasks, comprehension-understanding tasks, and opinion tasks. Snow then points out that within each task type there are different types of learning, such as memorization or drill and practice (see Ryan, 1981 and Kyllonen & Shute, 1989). For example, a traditional ATI study might contrast ability levels in a structured learning setting versus a self-guided learning environment. However, it is not likely that the traditional ATI study also considered the instructor's personality or other aspects of the learning environment that also affect learning outcomes. By presenting this possible taxonomy, Snow (1994) emphasizes the interactions possible within the treatment variable in addition to the interactions possible when aptitude and treatment are considered jointly.

The following literature review will summarize ATI studies based on student level, the design of the study, the purpose of the study (e.g., to examine ATI's?), how the dependent measure was developed, the type of aptitude and how it was measured, the treatment level, the instructional method, and the course content. Because this study is exploratory in nature and is being conducted as a literature review, no hypotheses will be formulated.

Method

Data Collection

A literature search was conducted using Eric, Psych Lit, and Social Science Citation Index to search for published articles and conference papers. Scientific and Technical Information Library Automation System (STILAS) was also searched for relevant Air Force technical reports. Finally, references from articles collected using the databases mentioned previously were collected for the purposes of this study.

While a number of studies were found via the literature searches, only a set of 23 studies were used for the purposes of this review. Studies were included that covered instructional settings in both schools and industry. Studies were omitted if the subjects under investigation were special needs students because there was concern that this subsample of subjects would not generalize to the Air Force population for which this review was intended. Only studies examining *aptitudes* were included in the review. In this review, aptitudes were restricted to cognitive, conative, and affective factors. Therefore, studies examining aptitudes defined as sex, race, or other individual differences not falling within the three categories mentioned above were excluded from this review. Finally, studies were excluded if they did not examine at least two different treatment levels in which at least a portion of all aptitude levels were exposed to all levels of the treatment. This inclusion rule was necessary because some studies were included in ATI reviews (and, therefore, obtained from the reference section) that, for example, investigated the effects of varying levels of an aptitude in one treatment. Although the studies themselves never claimed to examine aptitude treatment interactions, they were included in reviews of ATI's, which implied that they were considered ATI studies.

Data Analysis

Articles were coded according to student level, design of study (laboratory vs. field), how aptitude was defined and measured, how treatment levels were defined, the instructional methods used, the course content, and the dependent measures used (developed by researcher vs. previously established test vs. other). The unit of analysis was determined by the category being coded. In most studies it would be accurate to say that the study itself was the unit of measurement. However, there were cases where a study might be given more than one data point within a category. For example, it was not uncommon for a study to use more than one method for training subjects. In which case, the study would receive a data point for each method used. A study was not given additional points for each treatment condition, however.

Once all studies were coded within each category, the frequency of studies falling into each group was calculated. In addition, the percentage of the category in that group was also calculated.

Results

Table 1 presents frequencies for the level of student examined in the studies reviewed. As indicated, the largest percentage of subjects were undergraduates in college (35%), followed by upper elementary students (22%), high school students (13%), and others (13%). The other category represents subjects that graduated from high school and participated as subjects in laboratory experiments. This group can be differentiated from the trainee category who were undergoing training to learn job-related skills. The purpose of this breakdown was to explore the likelihood of generalizability to Air Force trainees. As three of the four most frequently represented categories were in the 15 to 25

age range, the likelihood of generalizability of these study results to the intended population is high because this is also the age range of many Air Force trainees.

Insert Table 1 about here

Table 2 shows that 54% of the studies were conducted in the field, while 33% were conducted in laboratories. Thirteen percent did not report where the study was collected. Again the generalizability of the studies reviewed in this paper is likely because the majority of the studies were conducted in the field where Air Force trainees are likely to undergo training.

Insert Table 2 about here

Table 3 reports the purpose of the study. These frequencies were calculated to determine whether researchers typically hypothesize aptitude treatment interactions or merely stumble across them while investigating other issues. Results indicate that most researchers make predictions about aptitude treatment interactions (91%).

Insert Table 3 about here

Because there was such a wide range of measures used for the dependent measure, three categories were formed. These included whether the measure was preestablished,

developed by the researcher, some other type of measure, or not reported. Preestablished measures included achievement tests, and reading and math tests. Measures developed by the researcher took the form of multiple choice, essay, or open ended questions and generally tested the subject's learning of the course material. Other types of measures included accuracy, number of errors, number of appropriate links, and number of concept words. Table 4 shows that dependent measures were most frequently developed by the researcher (52%). Twenty-six percent of the time dependent measures were measures other than paper and pencil tests and 17% of the time the measure was presetablished. One study did not report what type of dependent measure was used.

Insert Table 4 about here

The first four tables included some descriptive information about the types of studies included in this review. Table 5 begins the review of the aptitudes investigated in these studies. As mentioned previously, aptitudes can be broken into three categories including cognitive aptitudes, conative aptitudes, and affective aptitudes. In the studies reviewed, 71% investigated cognitive aptitudes, 16% investigated conative aptitudes, and 13% investigated affective aptitudes.

Insert Table 5 about here

Table 6 displays the types of cognitive aptitudes investigated by the studies reviewed. An attempt was made to put them in the framework discussed by Snow

(1994), however, two additional categories had to be created for aptitudes that did not fit in any of the other categories. The most frequently investigated aptitude was crystallized language (34%). These were followed by fluid reasoning (22%), visual perception (16%), and cognitive style (13%). In addition, general cognitive aptitude (9%), memory (3%), and motor skills (3%) were each investigated by a few studies. Auditory perception, speed, and idea generation, all cognitive factors discussed in Snow (1994), were not investigated by any of the studies included in this review.

Insert Table 6 about here

Table 7 presents the data points representing conative aptitudes investigated by studies under review. Only five data points examined conative aptitudes. Two data points looked at anxiety. Motivation, independence, and conformity were all investigated by one data point. The disparity between the number of data points associated with cognitive aptitudes (32) and conative aptitudes (5) may be due to the ambiguity associated with conative aptitude which is a difficult construct to define and measure.

Insert Table 7 about here

Finally, a total of seven data points obtained from the 23 studies examined affective aptitudes. These included attitudes/preferences (4) self-efficacy (2), and impulsivity (1). Again, this type of aptitude tends to be more unstable, and therefore,

difficult to measure and detect which may account for the lack of research attention directed toward affective aptitudes.

Insert Table 8 about here

Similar to the dependent measures, the type of measure used to assess aptitude was broken into three categories including, established measures, measures developed by the researcher, and other measures. Established measures included tests such as the Group Embedded Figures Test, the Scholastic Aptitude Test, and subtests from the California Assessment Test. The other category consisted of measures such as GPA and scores on an air combat maneuvering video game. Preestablished tests were used in 17 studies, 3 studies developed measures, and other measures were also used in 3 studies. The large disparity between the use of preestablished measures of aptitude and those measures developed by the researcher or other measures is likely related to the number of cognitive aptitudes examined. There are a number of standardized tests such as those mentioned above that are already designed to measure cognitive aptitude and are easy to administer in comparison to developing aptitude measures.

Insert Table 9 about here

The next several tables presented explore issues relating to the treatments used in the studies under review. In Table 10, the treatment levels varied a great deal, and thus, were consolidated based on similarity of the training levels for a more comprehensible

cognitive aptitudes examined. There are a number of standardized tests such as those mentioned above that are already designed to measure cognitive aptitude and are easy to administer in comparison to developing aptitude measures.

Insert Table 9 about here

The next several tables presented explore issues relating to the treatments used in the studies under review. In Table 10, the treatment levels varied a great deal, and thus, were consolidated based on similarity of the training levels for a more comprehensible review. The total number of data points was 24. Of this total, 42% of the treatments compared structured training to unstructured training, 21% compared elaborative to abstract training, and 13% compared a graphical presentation to a textual presentation. Each of the remaining treatments contributed 1 data point to the total. These treatments include, whole vs. segmented tasks, participation vs. no participation, cognitive modeling vs. lecture, limited vs. unlimited computer access, small vs. large group learning, and form discrimination training vs. cognitive modeling training.

Insert Table 10 about here

Table 11 displays the frequency of data points for the instructional methods used in the studies reviewed. Lecture accounted for 21% of the data points, practice accounted for 18% of the data points, while workbook/textbook learning and training via computer each accounted for 15% of the total data points. Video training, training using graphs,

and training in a laboratory each accounted for less than 10% of the total data points. Finally, 9% of the data points were relegated to the "did not report" category because that information was not provided by the researcher. These results are consistent with what one would expect, indicating that traditional instructional techniques such as, lecture, practice, and textbook learning are still very prevalent. Interestingly, five data points came from training using a computer. As we move further into the information age, this method of training will undoubtedly become more prevalent.

Insert Table 11 about here

Table 12 presents the course content covered during training sessions in the studies. Science was most frequently taught in studies (3), followed by math, flight engineering, electricity, psychology, and software use/computer programming which each contributed two data points. One data point was contributed by each of the other treatment levels: learning, general studies, electronic mail system, paper folding, Malay-English word pairs, information seeking, cognitive abilities, and idea generation. Finally, two studies did not report the content learned during instruction. As evidenced from the diversity of the training content, some studies discussed more general topics in training while other studies focused training on specific learning tasks. None of the content areas, however, were particularly well represented across the studies.

Insert Table 12 about here

The last set of tables pertain to the significance of the interactions. For these analyses, studies were again separated according to the type of aptitude examined (e.g., cognitive, conative, or affective). Table 13 presents the frequencies for data points finding either an aptitude treatment interaction or no interaction when investigating cognitive aptitude. The results indicate that of the 37 data points, 24 found interactions and 13 did not. While no conclusions can be drawn as to whether there are ATI's for cognitive aptitude, the results of this analysis indicate that more studies found evidence that ATI's do in fact exist.

Insert table 13 about here

Table 14 presents the results for ATI's examining conative aptitude. In this case, all three of the data points found significant aptitude treatment interactions. Table 15 indicates that 7 data points were associated with affective aptitude and of these 2 found significant results and 5 did not find evidence of affect by treatment interaction. Finally, several studies examined more than one aptitude and considered them in a three-way interaction. In each case, the type of aptitudes were mixed (e.g., cognitive x affective x treatment), therefore, a separate analysis was conducted for these specific cases. The analysis, presented in Table 16, indicate that each of the 4 data points found significant interactions.

Insert Table 14 about here

Insert Table 15 about here

Insert Table 16 about here

Discussion

The first three tables provide some information about how ATI studies have been investigated to date. First, the typical study uses subjects that range in education level from upper elementary to undergraduate college level. There are relatively few studies involving trainees being instructed in work-related skills. Second, more studies are conducted in the field than in laboratories. While the realistic setting of field studies can be an advantage, Shute (1993) points out that field studies are limited in the ability to control for extraneous factors and, therefore, detect aptitude treatment interactions. Third, most studies are, in fact, conducted to examine ATI's and are, thus, designed methodologically for that purpose.

The review of the literature studying ATI's also indicates that most researchers develop their own dependent measures. While this does not have to be a weakness of a study, I noticed that some researchers failed to describe how the measure was developed

(e.g., Shaw & Bunt, 1979). Some omitted any specific discussion about the content of the measure (e.g., Ysseldyke, 1977). And, very few researchers included any information regarding the psychometric properties of the measure, let alone subjecting the measure to pilot testing prior to using it for research purposes (e.g., Beukhof, 1986).

Tables 5 through 9 discussed research trends regarding aptitudes. Several authors have noted that aptitudes can be divided into three categories: cognitive, conative, and affective (Shute, Lajoie, & Gluck, in press; Snow, 1989). The review indicates that researchers study all three types of aptitudes. There is, however, much more research attention devoted to the study of cognitive abilities. As mentioned previously, the stability of cognitive aptitudes relative to either conative or affective aptitudes makes them easier to measure. In addition, cognitive aptitudes have been found to predict performance better than other types of aptitude which may also contribute to the relative prevalence of studying cognitive aptitude.

A breakdown of the types of cognitive aptitudes studied reveals that crystallized language, fluid reasoning, spatial perception, and cognitive style are most frequently studied. In comparison, no one type of conative or affective aptitude was studied more frequently than the others.

Finally, the review showed that contrary to the results found for how dependent measures are obtained, researchers rely more heavily on preestablished measures of aptitude. It was still common to omit any report of the psychometric properties of the measures chosen. However, of the 3 studies that developed their own measures of aptitude, 2 were subjected to pilot testing prior to use in research.

Tables 10 through 12 summarize the trends concerning aspects of the treatment. For example, Table 10 reveals that the most frequently used treatments were a structured vs. unstructured approach and an elaborated vs. abstract approach. These two treatment levels are similar in that one level provides the student with supplemental information while the other level provides relatively little information. More specifically, a structured approach provides more direction to the student while the unstructured approach is more self-guided. Similarly, the elaborated approach provides more information about the learning task than the abstract approach.

The review of ATI literature shows that the traditional methods of instruction such as lecture, practice, and use of a textbook or workbook were the most frequently used methods for conveying training material. Interestingly, there were 5 studies that conducted training using a computer. While training on computers was necessary to learn some of the tasks (e.g., programming skills), it was not required for all of tasks.

Finally, a review of the content taught in the studies revealed that a wide range of topics were covered in these studies. Surprisingly, the most frequently taught topic was science and that was only taught in 3 of the 23 studies.

The last set of tables summarizes the results of the ATI studies reviewed. Over the years many researchers have debated whether aptitude treatment interactions exist. Two meta-analyses came to the conclusion that they do not based on their results (Bennett, 1996; Kavale & Forness, 1987). Although this review does not calculate a mean effect size across the studies to determine the magnitude of the effect of ATI's, it does indicate that of the studies reviewed 65% of the studies investigating cognitive aptitude, 100% of the studies investigating conative aptitude, and 29% of the studies

investigating affective aptitude indicated finding ATI's. In addition, of the studies examining more than one type of aptitude, 100% found ATI's. This is compelling support for the existence of ATI's, however, these results must be interpreted with caution. There was some tendency to emphasize significant results and gloss over nonsignificant results which could have lead to some bias in how the results were included in this study (Beukhof, 1986).

No frequency analyses were conducted to examine trends in the aptitude treatment interactions that would make it possible to make specific predictions regarding which aptitude groups learn best under specific conditions. Unfortunately, due to the diversity of the types of aptitudes and treatments, it was determined that so few data points could be combined that a frequency analysis would give little additional information. However, it is possible to comment on some trends in the literature. Specifically, where cognitive aptitudes were considered, it appeared that high aptitude students tended to perform better in unstructured learning environments while low aptitude students tended to perform better in highly structured learning environments. Also, low cognitive ability students benefitted from elaboration while high cognitive ability students did not.

Conclusions and Recommendations

This review highlights three types of aptitudes that can be considered in developing a taxonomy of learning characteristics for training. The review suggests that while it may be more difficult to measure conative and affective aptitudes, they can interact with aspects of training to influence learning outcomes. The review also suggests one framework for considering the numerous cognitive aptitudes namely a higher order factor structure. One additional framework for considering cognitive ability should also

be mentioned. Snow and Lohman (1984) examined correlations among a number of ability tests. They found that the tests that correlated highly with general intelligence were located toward the center of a radex and those correlating less highly were located toward the periphery. They also pointed out that those tests located toward the center were more highly related to complex tasks while those tests located in the periphery were more related to simple tasks. Finally, the tests formed factor clusters with verbal or crystallized abilities, abstract-reasoning or fluid-analytic abilities, and spatial-visualization abilities nearest to the center. Thus, when examining cognitive aptitudes it is necessary not only to consider the level of specificity at which one measures cognitive ability, but also the tasks performed in training.

The review also suggests that it is relevant to consider several aspects of the learning environment or treatment in addition to aptitudes. First, a taxonomy should consider the treatment differences. This review suggests that training programs may differ in the amount of structure given to the students, the amount of elaboration about the course content, whether the material is presented graphically, textually, or verbally, and whether training is computer based or in traditional lecture format. The review also indicates that it is relevant for a taxonomy to consider how the material is presented (e.g., lecture, practice, textbook, video, computer). Finally, a taxonomy should consider course content. I should mention that there was some suggestion that I should code for whether feedback was given to trainees and how it was delivered. While I did not code for these variables, I did look for them. I found no mention of feedback given to any of the trainees in the literature reviewed in this study.

Assessment of the Reliability of Ground-Based Observers for the detection of Aircraft

Jason McCarley
Graduate Research Assistant
Department of Psychology

University of Louisville
2301 South Third Street
Louisville, Ky 40208-1802

Final Report for:
Graduate Student Research Program
Armstrong Laboratory

Sponsored by:
Air force Office of Scientific Research
Bolling Air Force Base, DC

and

Armstrong Laboratory

September 1996

Assessment of the Reliability of Ground-Based Observers for the Detection of Aircraft

Marc L. Carter
Assistant Professor
Department of Psychology
University of South Florida

Jason McCarley

Department of Psychology
University of Louisville

Abstract

In situations in which ground-based lasers are propagated through the atmosphere, either for entertainment or scientific pursuits, there is the chance that aircrew may be exposed to the beam. In most cases this exposure would not be eye-hazardous, but the effects of flashblindness and veiling glare can nonetheless impair mission performance, with potentially catastrophic consequences. In most situations where such lasers are employed, ground-based observers attempt to identify aircraft that are in or near the beam path; occasionally these observers are aided by FAA radar feeds that can assist them in locating these aircraft. In this study we attempt to determine the effectiveness of observers in the detection of aircraft under a variety of conditions, including day versus night, and with and without the assistance of a radar feed. Preliminary data collected at Sandia National Labs in Albuquerque, NM, suggest several points. First, detection range is very much greater at night than in the day, probably due to the high contrast between the aircraft and night sky from aircraft lighting, and the increased visual sensitivity of the observers in scotopic viewing. Second, the assistance of a radar feed for daytime observation is important in aircraft detection, not so much to increase the range at which the aircraft is visually acquired, but to increase the likelihood that the aircraft will be detected at all. In further analysis of the complete data set we will examine the impact of various ground and sky conditions that can mitigate the performance of the observers.

Assessment of the Reliability of Ground-Based Observers for the Detection of Aircraft

Marc Carter and Jason McCarley

As lasers have come into widespread military application, enormous effort has been dedicated to ensuring their safe use. Much of this has aimed at assessing the biological effects of ocular exposure to laser, and at determining eye-safe levels of exposure for a variety of lasers, both battlefield and other. It is well established, though, that levels of exposure too weak to cause biological damage can impair visual function (Thomas, 1994). Either glare--scattered light veiling a scene (Pulling, Wolf, Sturgis, Vallancourt, & Dolliver, 1980; Vos, 1984)--or flashblindness--the loss of sensitivity which follows exposure to intense light (Brown, 1965)--can cripple visual performance. Both effects are temporary, but consequential. For pilots, momentary functional blindness can devastate, most importantly at mission-critical junctures, when the aircraft is more likely to be near the ground (targeting, landing and takeoff).

For some time glare and flashblindness have been of great interest to the military, increasingly as the potential for laser eye exposure has increased due to the potential for lasers to be wielded as weapons by enemies or the actual use of lasers as tools by friends. More recently, however, laser exposure to civilian aircrew has become a concern as the number of lasers employed in entertainment, advertising, and scientific applications has increased (Aviation Week & Space Tech, Sept. 26, 1994; Weiss, 1996). The potential for danger became a reality when a commercial airline pilot, upon departure from Las Vegas' McCarran International Airport, was temporarily blinded by a casino laser, and forced to turn control of his aircraft over to the copilot (Scott, 1995). This event, having been preceded by complaints of approximately fifty similar

incidents, prompted the FDA to temporarily prohibit laser light shows in the city of Las Vegas, and to consider a similar, nationwide moratorium (Air Line Pilot, 1996). But such a prohibition could be fully effective only if it also banned use of atmospherically transmitted scientific lasers, whose high-intensity, low-divergence beams remain hazardous over greater distances than the typically low-power beams employed by commercial displays (Weiss, 1996). A summary proscription of outdoor lasers, then, would confiscate a tool of both science and commerce. This is a cost to be incurred if necessary, but to be avoided otherwise.

One popular safeguard against the dangers of outdoor lasers is the simple use of ground-based observers to monitor the airspace through which the lasers are deployed, and to control the firing of lasers when there are aircraft in or approaching the beam path. A more sophisticated and potentially more reliable operation employs a radar feed from (typically) FAA radar located at an airport in the vicinity. Such a system, although perhaps better than simple observation, still suffers from several defects, among them the inability of FAA radar to accurately locate aircraft near the ground (where they most obviously are during takeoff and landing, perhaps the most critical portions of any flight) as well as in the "blind spot" inherent in ground-based radar systems (Weiss, 1996), and the temporal lag involved in registration of an aircraft's location on the radar display. The combination of a radar feed along with visual observation offers the best solution present to date, in that a radar operator can direct the observers to a region of airspace in which an aircraft is indicated by the radar; the visual observers then are able to very accurately determine the current location of the aircraft with respect to the beam path.

Under most current circumstances, however, unassisted observers are likely to be the sole safeguard against laser exposure to commercial and private aircraft, in spite of the fact that their

reliability has not been assessed. Past research has generally examined the ability of observers to acquire aerial targets appearing against the relatively uniform backdrop of a daylight sky (e.g., Akerman & Kinzly, 1979; O'Neal, Armstrong, & Miller, 1988), and models of visual target acquisition have likewise, explicitly or implicitly, assumed photopic illumination (e.g., Akerman & Kinzly, 1979; Koopman, 1986). For several reasons, the results of these efforts cannot be easily generalized to many of the conditions under which outdoor lasers are deployed. Under photopic illumination, a distant aircraft at threshold contrast will likely appear as a still, dark patch, undistinguished by color or motion (Akerman & Kinzly, 1979; Koopman, 1986). On the other hand, color, motion, and flicker might be the most salient source of information about the approach of a night-flying aircraft bearing an array of identification, landing, and stroboscopic anticollision lights. This information, along with increased visual sensitivity that accompanies a decline in ambient illuminance, may well allow acquisition ranges measured at night to exceed those obtained in daylight. Unfortunately, the dark adapted eye is most susceptible to glare (Sturgis & Osgood, 1982) and nighttime performance, vulnerable to the ambient light of a city or airport, may be volatile.

Daytime and nighttime target acquisition are likely to represent fundamentally different visual abilities, such that performance between tasks will differ substantially. The present research will therefore examine the ability of visually unaided ground-based observers to monitor airspace under photopic and scotopic conditions, and with consideration for the terrain over which search is conducted. Additionally, acquisition ranges will be compared between conditions in which observers work in concert with a radar operator, and those in which observers freely scan an assigned area of space. Results will be used to assess the efficacy of such observers as a safeguard

against exposure of aircrew to atmospherically propagated lasers.

Method

Observers

Eight male volunteers were recruited from among the employees of Sandia National Laboratory, and paid for their participation. The observers were screened to ensure normal color vision, and visual acuity of 20/20 or better, and ranged in age from 33 to 35 years. The observers used in this study had a variety of experience at spotting for aircraft, some having never performed the task prior to testing, and others having 30 or more hours experience.

Procedure

Data were (and continue to be) collected at Sandia National Laboratories' LAZAP Facility, located on the Sandia Military Reservation, Albuquerque, New Mexico. The facility is the site of a high-powered, low beam-divergence, ruby laser, used for the calibration of on-orbit DOD and DOE satellite payloads. The data are being collected apart from the Facility's normal operations, and do not involve any use of lasers.

Subjects will be stationed approximately 3 m above ground level, on a rooftop near the Laboratory's outdoor laser, during times at which the laser is not operational. For day observation sessions, subjects are encouraged to use appropriate sunscreen. Test sessions last approximately four hours, and are divided into one-hour shifts of data collection separated by periods of rest, with a long rest (approximately one-half hour) after two hours.

A simple 2 (unaided vs. aided search) by 2 (photopic vs. scotopic illumination) factorial

design will be used to analyze data. In all conditions, observers are asked to indicate when they visually acquire the target. In the free scanning (unassisted) conditions, the airspace surrounding the point of observation is divided into hemifields facing either north and south or east and west. Throughout a single shift, observers monitor an assigned hemifield of airspace, and report any aircraft they acquire within that airspace. In aided scanning conditions, observers monitor 360 degrees of airspace each shift, but are given the range, azimuth, and altitude, of targets entering the monitored airspace. Daylight shifts are run between 10:00 AM and 2:00 PM, in order to minimize cues from glint off the aircraft or having the aircraft in line with the position of the sun.

In the assisted condition, aircraft location information is provided by a feed from the FAA radar at the Albuquerque airport, and are noted by a Radar Airspace Monitoring System (RAMS) operator. Data collected include the time of day at which each target appears, cloud cover and visibility at that time, the range, azimuth, and elevation of the target when notification is made, and then the range, azimuth and elevation of the target when it is detected. If a target goes undetected, its nearest approach to the observer is recorded. The straight-line distance from the observer at which a target is acquired is the most immediate measure of performance.

Preliminary Results, and Some Discussion

At the writing of this report, data collection was not complete. However, data from approximately half of the observers is available. For this preliminary analysis, we use "clean" data; that is, aircraft detection that relied on anything other than simple detection was eliminated from the summary. Hence, no detections that were aided by glint, contrails, or sound, or detections

that were hampered by clouds, veiling glare from nearby lights, or obscured by buildings, are included. We also eliminated data from aircraft that were near take-off or landing, although in the night conditions detection via landing lights from distant aircraft is allowed. The data are also presented as linear miles from the observer, without consideration of azimuth or elevation; those parameters will be considered in the complete analysis. The data are classified as "hits," which is the point at which the observer reported spotting the aircraft, or "misses," in which case the data reflect the aircraft's closest point of approach to the observer.

First, the data from the day conditions:

	Assisted			Unassisted		
	Mean	SEM	N	Mean	SEM	N
Hits	11.37	.681	63	8.03	.140	28
Misses	13.07	.630	16	12.11	.313	21

There are a couple of points of relevance in these data. First, note that observers were detecting aircraft at greater distances in the assisted rather than unassisted conditions. This clearly demonstrates that for daylight observations the assistance of the radar feed is to be preferred. Another aspect of these data is the difference in the quantity of detections between the two conditions, but no firm conclusions can be drawn from this since, although we attempted to equate the amount of time the observers spent in each condition, we obviously had no control over the air traffic during those sessions. It is possible, however, to remark on the difference in *proportion* of aircraft detected versus missed. Of the 79 aircraft present during the assisted sessions, the observers spotted very nearly 80%, whereas in the unassisted conditions they only

spotted 57%. This also lends support for the benefit of having assistance in searching for aircraft.

Next, the data from the night observations:

	Assisted			Unassisted		
	Mean	SEM	N	Mean	SEM	N
Hits	23.49	.147	49	24.60	.281	38
Misses	22.72	2.557	3	21.00	1.38	6

The first thing to note about these data is that there is only a small (but nonetheless reliable) difference in mean detection range for the two conditions (assisted versus unassisted), which at first may seem surprising. However, this is almost surely due to the much greater ability of the observers to detect aircraft by means of lights (either anticollision or landing lights) at night, when the observers sensitivity is greater and the aircraft's contrast against the sky is high, regardless of whether or not the observer is directed to the area of sky in which the aircraft can be found. They are simply much easier to see at night, and the RAMS feed is of much less import. Also note that the proportions of hits versus misses is much the same for the two conditions, although there is a slightly higher proportion in the unassisted condition. The last, and perhaps most salient, point is to compare the data from the day and night sessions. It is clear that the effective range of the observer is very much greater (at least twice as great) at night than in the daytime. Although not surprising, it is important to note: observers simply see the aircraft farther away at night, due to the presence of landing lights or anticollision strobes. When the contrast between the aircraft and the background is lower, such as in the daylight conditions, performance suffers.

To sum up, we have learned from the preliminary examination of the data that the RAMS operator is very important during the daylight, both to increase the range at which aircraft may be detected, and to increase the likelihood that an aircraft will be detected. At night, however, this is much less important: observers routinely detected aircraft that were at times outside the range of the radar information, and much less frequently failed to note the presence of an aircraft in the monitored region of airspace. We look forward to completing data collection and generating a more thorough analysis of these data, including consideration of azimuth (for city lights, mountains, sun position, and the like) and elevation above the horizon.

References

Akerman A, Kinzly RE. (1979). Predicting aircraft detectability. *Human Factors, 21,* 277-291.

Brown JL. (1965). Flashblindness. *American Journal of Ophthalmology. 1965; 60,* 505-520.

Hood DC, & Finkelstein MA. (1986). Sensitivity to light. In Boff, KR, Kaufman, L, & Thomas, JP (Eds.), *Handbook of Perception and human performance.* New York: Wiley.

Editors. (1996). FAA, FDA put hold on outdoor lasers, but only at Las Vegas. *Air Line Pilot, 65,* 40-41.

O'Neal MR, Armstrong HG, Miller, RE. (1988). Further investigation of contrast sensitivity and visual acuity in pilot detection of aircraft. Harry G. Armstrong Aerospace Medical Research Laboratory, Wright-Patterson AFB, OH: 1988; AAMRL-TR-88002.

Pulling, NH, Wolf, E, Vaillancourt, DR, & Dolliver, JJ (1980). Headlight glare resistance and driver age. *Human Factors, 22,* 103-112.

Olzak LA, Thomas JP. Seeing spatial patterns. In Boff, KR, Kaufman, L, & Thomas, JP (Eds.), *Handbook of Perception and human performance.* New York: Wiley.

Scott WB. (1995). Southwest pilot injured by laser. *Aviation Week & Space Technology,* Nov 20, 1995.

Sturgis, SP, & Osgood, DJ (1982). Effects of glare and background luminance on visual acuity and contrast sensitivity: Implications of driver night vision testing. *Human Factors, 24,* 347-360.

Thomas SR. (1994). Review of personnel susceptibility to lasers: Simulation in SIMNET-D for CTAS-2.0. Occupational and Environmental Health Directorate, Optical Radiation Division, Brooks AFB, TX: 1994; AL/OEO-TR-1994-0060.

Vos JJ. (1984). Disability glare--a state-of-the-art report. *C.I.E Journal, 3,* 39-53.

Weiss SA. (1996). Aviation officials look warily at scientific lasers. *Photonics Spectra,* May.

A QUANTITATIVE EVALUATION OF AN INSTRUCTIONAL DESIGN SUPPORT SYSTEM: ASSESSING THE STRUCTURAL KNOWLEDGE AND RESULTING CURRICULA OF EXPERT AND NOVICE INSTRUCTIONAL DESIGNERS

Theresa L. McNelly
Graduate Student
Department of Psychology

Texas A&M University
College Station, TX 77843-4235

Final Report for
Graduate Student Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Brooks Air Force Base, TX

and

Armstrong Laboratory

September 1996

A QUANTITATIVE EVALUATION OF AN INSTRUCTIONAL DESIGN SUPPORT SYSTEM: ASSESSING THE STRUCTURAL KNOWLEDGE AND RESULTING CURRICULA OF EXPERT AND NOVICE INSTRUCTIONAL DESIGNERS

Theresa L. McNelly
Graduate Student
Department of Psychology
Texas A&M University

Abstract

The Guided Approach to Instructional Design Advising (GAIDA) is an automated instructional design (ID) tool developed at Armstrong Laboratory's Technical Training Research Division. Based on Gagné's nine events of instruction, GAIDA was developed to aid content domain experts, who are novices in instructional design, in the planning, development, and implementation of quality computer-based instruction (CBI). A systematic, quantitative evaluation is being conducted to determine whether GAIDA can be used to acquire the skills that would otherwise be obtained by means of long-term, on-the-job, instructional design experience. Reaction, learning, and behavioral measures will be collected. Learning will be assessed by administering traditional paper-and-pencil knowledge tests and by investigating the structural knowledge of the participants. A comparison of the knowledge structure of novice instructional designers and expert instructional designers will be conducted before and after the implementation of GAIDA. Several knowledge representation techniques (namely, multidimensional scaling and Pathfinder) will be used to assess the underlying mental models of novice and expert instructional designers to determine (1) the similarity of novice and expert mental models before the implementation of GAIDA, and (2) whether the implementation of GAIDA results in an increase in similarity between the expert and novice mental models of instructional design. Behavior will be assessed by obtaining courseware samples from the participants, which will be rated as to the extent that the courseware incorporates the nine events of instruction proposed by Gagné.

A QUANTITATIVE EVALUATION OF AN INSTRUCTIONAL DESIGN SUPPORT SYSTEM: ASSESSING THE STRUCTURAL KNOWLEDGE AND RESULTING CURRICULA OF EXPERT AND NOVICE INSTRUCTIONAL DESIGNERS

Theresa L. McNelly

Introduction

Subject matter experts (SMEs) are often used to design curricula and develop technical training courses for their area of specialized knowledge. The typical SME is therefore knowledgeable about a particular content domain, but not necessarily versed in the development of instructional design. The Guided Approach to Instructional Design Advising (GAIDA)[1] is an instructional design support system developed at Armstrong Laboratory, Brooks Air Force Base, in San Antonio, Texas. It is a CD-ROM based multi-media software program designed to aid novice and expert instructional designers in the development of quality computer-based instruction (CBI). GAIDA is based on the nine events of instruction proposed by Gagné (1985) to be essential in the instructional design process. These nine events are: gaining attention, informing the learner of the objectives, stimulating recall of prerequisite knowledge, presenting the stimulus material, providing learning guidance, eliciting the performance, providing feedback about performance correctness, assessing the performance, and enhancing retention and transfer. GAIDA presents step-by-step guidance in applying Gagné's nine events of instructional design.

GAIDA operates in two modes: lesson and guidance. In the lesson mode, the user can peruse an assortment of interactive courseware examples. This lesson library or casebase

[1] The Guided Approach to Instructional Design Advising (GAIDA) is Armstrong Laboratory's research version of the instructional design support software. GAIDA has evolved extensively since it's original creation and is currently in the process of being marketed as the Guide to Understanding Instructional Design Expertise (GUIDE). The two versions are, for the most part, identical.

demonstrates computer-based instruction for a variety of training objectives and presentation techniques. In the guidance mode, an explanation of each of the nine events of instruction is presented, as well as demonstrations of how to incorporate each of the nine events effectively to create meaningful interactive courseware. The two modes work in conjunction with each other, allowing the user to jump from one mode to the other (Gettman & Whitehead, 1995). This provides instructional designers with complete control over the organization of the sequence of learning throughout the program. Further, GAIDA is equipped with an on-line note taking feature that allows the user to record thoughts and ideas while interacting with the software. It is important to note, however, that GAIDA does not actually author the CBI. The intention of GAIDA is to aid in the development of the CBI content through the introduction of applicable instructional design practices.

Several formative evaluations of GAIDA have been conducted (Gagné, 1992; Tennyson & Gettman, 1995) and the initial results have been overwhelmingly positive (Gettman, 1995). An initial evaluation of GAIDA's general approach was conducted (Gagné, 1992). This initial evaluation used a model of formative evaluation provided by Dick and Carey (1991). This model considers three main criteria: (1) Clarity: are the directions clear?; (2) Impact: what is the effect of the instruction on the achievement of the objectives?; and (3) Feasibility: given certain support and time allocation, how feasible is the instruction? The participants were given general instructions and guidance and asked to design instruction following the format outlined by GAIDA. Findings indicated that (1) students experienced no difficulty in understanding or using the instruction of the GAIDA lesson, (2) the lessons

they designed followed closely the model task provided as an example, and (3) 83% of the students judged the lesson they designed to be useful and effective.

In a more extensive evaluation, Tennyson and Gettman (1995) used instructional technology graduate students enrolled in an instructional design course over several semesters to assess attitudes toward and perceptions of GAIDA, as well as assess student learning. Several findings resulting from this evaluation are noteworthy. First, participants rated the quality of GAIDA fairly high. The quality of the texts and graphics received excellent ratings. Further, participants reacted positively to GAIDA's ability to give the user control of the sequence of learning within the program. Second, participant attitudes about GAIDA's ability to aid in designing instruction shifted from indifference, at the beginning of the course, to positive, by the end of the course. Finally, results indicate that the use of GAIDA was somewhat more effective than a traditional textbook version in presenting Gagné's nine events of instruction. Using the grade received on end-of-course projects as a criterion, participants who opted to learn the nine events using GAIDA technology received higher grades on the project than students who chose to learn the nine events from a textbook.

These two initial evaluations were highly favorable of GAIDA. Other evaluations have not been as positive, however. Asiu (1995) conducted focus groups with novice Air Force CBI developers and instructional technology graduate students using GUIDE (Guide to Understanding Instructional Design Expertise). Responses from the group of Air Force developers indicated that GUIDE did not always provide a coherent link between the lesson objectives represented by the example lessons and content areas of interest to the Air Force

users. For example, Air Force developers using GUIDE had difficulty understanding how the program could aid them in designing and organizing CBI when the case examples/samples did not closely resemble their area of expertise. Results from the second focus group activity, the instructional technology graduate students, reflected similar findings. Additionally, the second focus group identified some general notions about GUIDE not previously recognized. For example, those who had more computer experience, especially with Windows-based programs, had higher perceptions of the use and utility of GUIDE. To date, the evaluations of GAIDA/GUIDE have been qualitative in nature and the results have been somewhat mixed. The present study will evaluate GUIDE in a more systematic, quantitative manner.

The effectiveness of a training program can be evaluated in terms of reaction, learning, behavioral, and/or results criteria (Kirkpatrick, 1987). Reaction criteria measure how well the participants or trainees liked the program, including its content, the trainer, the methods used, and the training environment. Learning criteria assess the knowledge gained by the trainees. Knowledge acquisition is typically measured by paper-and-pencil tests (Wexley & Latham, 1991). Collecting behavioral criteria measures addresses the need for assessing changes in a trainee's overt behavior upon returning to the job, in addition to the knowledge that was acquired during training. Finally, results criteria are collected in order to assess the actual benefit, at an organizational level, of the training program. Typical results criteria include a reduction in turnover, increase in quality and quantity of goods produced, increase in sales, or a reduction in accidents.

This study proposes to assess GUIDE through the evaluation of reaction, learning, and behavioral criterion measures. First, reactions of instructional designers will be assessed

after being exposed to GUIDE for a specified period of time. Previous evaluations of GAIDA (Gagné, 199 ; Tennyson and Gettman, 1995) have indicated that reactions have been overwhelmingly positive. Therefore, in alignment with past research, it is hypothesized participants will generally have favorable attitudes towards GUIDE. Specifically, it is hypothesized that they will report that GUIDE is easy to understand and use when developing CBI.

Second, learning of instructional design in general and Gagné's nine events in particular will be assessed by administering traditional paper-and-pencil knowledge tests before and after the implementation of GUIDE. It is hypothesized that there will be a significant increase in test scores between the pre-training administration and the post-training administration.

In addition to traditional measures of learning, the structural knowledge of the participants will be assessed before and after the implementation of GUIDE. A number of researchers have become interested in measuring how knowledge is organized in memory by studying knowledge structures or mental models (Dorsey & Foster, 1996). This interest reflects the recognition that the *organization* of knowledge stored in memory is perhaps of equal or greater importance to the *amount* of knowledge stored in memory (Johnson-Laird, 1983; Kraiger, Ford, & Salas, 1993; Rouse & Morris, 1986). Mental models have been defined as a "rich and elaborative structure, reflecting the user's understanding of what the system contains, how it works, and why it works that way. It can be conceived as knowledge about the system sufficient to permit the user to try out actions mentally before choosing one to execute." (Carroll & Olson, 1988, p. 51).

There are two important characteristics of mental models: the type or complexity of the stored elements in memory and the organization or interrelationships among the model elements. Numerous studies of expert/novice differences (i.e., Adelson, 1981; Chase & Simon, 1973; Chi, Feltovich & Glaser, 1981; McKeithen, Reitman, Rueter & Hirtle, 1981) suggest that experts and novices store elements of knowledge differently in a given content domain. Results of these studies indicate that novices create different mental models for defining the problem and deriving a solution. Experts, on the other hand, have far more complex knowledge structures that contain elements for both problem definition and solution strategies (Glaser & Chi, 1989). Thus, experts enjoy the advantage of having quick access to solution strategies once the problem has been identified because that strategy is closely linked in memory to the problem node. In contrast, novices are much slower and more awkward at deriving solutions because the problem nodes and solution nodes are farther apart in memory and separate searches must be undertaken in order to solve just one problem (Schvaneveldt, Durso, Goldsmith, Breen, Cooke, Tucker, & De Maio, 1985). By collecting information on both the amount of learning (traditional paper-and-pencil test) and the organization of learning (assessment of mental models), the convergence between these two metrics of learning, and the relationship between the organization of knowledge and training outcomes can be determined.

There are three distinct steps involved in assessing structural knowledge: knowledge elicitation, knowledge representation, and comparing and contrasting an individual's knowledge representation to some standard (i.e., instructor's knowledge representation; Goldsmith, Johnson, & Acton, 1991). Knowledge elicitation, the first step, assesses an

individual trainee's understanding of the relationships among a set of concepts central to the material contained in the training program. A variety of methods have been used to elicit these relationships. These include word associations, ordered recall, card sorting, and numerically rating the degree of interrelatedness (or similarity) between each pair of concepts deemed to be highly related to the content domain of interest. These procedures or techniques typically produce a matrix of proximity values. Proximity values are the ratings given by the subject as to the similarity between a single pair of concepts.

The second step, knowledge representation, involves structuring the data into some descriptive representation of the concepts, usually using one or more of several scaling techniques available. The most frequently used scaling techniques are multidimensional scaling (MDS), cluster analysis, additive trees, and networks. The third step involves the comparison of the individual's derived knowledge representation to some standard, usually an expert in the content domain of interest (Goldsmith et al., 1991).

In the proposed study, knowledge elicitation will be accomplished by obtaining proximity values for each pair of concepts from each individual trainee. These proximity values will then be analyzed by using multidimensional scaling and network scaling techniques to represent the knowledge structure for each novice instructional design trainee. These knowledge structures will then be compared to expert instructional designers to determine the degree of similarity between the expert and novice knowledge structures before and after the implementation of GUIDE. It is hypothesized that there will be an increase in similarity between the novice and expert instructional designers after the implementation of GUIDE. Further, it is hypothesized that there will be high variability across the mental

models derived from the novice instructional designers at time one, and that this variability across the novice mental models will decrease at time two.

Finally, behavior measures will be collected before and after the implementation of GUIDE. Behavior, in this case, will be displayed in the extent to which Gagné's nine events of instruction are incorporated in the curricula developed after interacting with GUIDE for a specified period of time. It is hypothesized that there will be an increase in the degree to which Gagné's nine events are incorporated in the CBI developed before and after interacting with GUIDE.

Method

Participants

There are three potential test beds available to obtain the expert and novice instructional designers needed for the proposed study. The first test bed is comprised of approximately 15 courseware designers at each of the four Air Force Technical Training Centers. The second test bed consists of over 20 CBI designers from a training facility for Food Safety and Inspection Services, a division of the U. S. Department of Agriculture. The third potential test bed is comprised of approximately 170 instructors who work for the Texas Engineering Extension Service (TEEX) at Texas A&M University in College Station, Texas. All three test beds are directly involved in the development, design, and implementation of computer-based instruction. Although the exact number of participants cannot be ascertained at the current time, an adequate sample of expert and novice instructional designers will be obtained to permit the analyses proposed in this study. Expert instructional designers will be differentiated from novice instructional designers by determining the length of time each

instructor has been directly involved in developing courseware. Those instructors with significantly more on-the-job experience in instructional design will be categorized as experts.

Measures

Reactions - Trainee's affective response to the training will be assessed by collecting traditional end-of-course reaction evaluations. Items included in the reaction measure will assess the degree to which trainee's feel that GUIDE's instructions are clear, comprehensible and workable, the extent to which they feel that GUIDE assists them in developing CBI, the extent to which they intent to incorporate the principles outlined by GUIDE in subsequent CBI they develop, and the extent to which they will recommend GUIDE to other peers and colleagues involved in developing CBI.

Learning - Learning will be assessed by administering a traditional paper-and-pencil knowledge measure before and after the implementation of GUIDE. The items on the measures will cover the content of GUIDE to assess the extent that the knowledge of the nine events has been acquired by the trainee.

Learning will also be assessed by investigating the participants' mental models of instructional design. The knowledge domain for this study is instructional design in general and Gagné's nine events of instruction, in particular. When devising a list of concepts particular to a specific content domain, it is advisable to limit the number of items in the set to simplify data collection and interpretation of results (Cooke & McDonald, 1987). By limiting the number of concepts to be analyzed, the researcher is assured that only relevant concepts are being investigated. A comprehensive list of concepts relevant to instructional

design and Gagné's nine events will be obtained. Subject matter experts (instructional designers, in this case) will then rate each of the concepts in terms of importance in the design of CBI. The concepts with the highest overall importance ratings will be used in this study. A list of potential concepts can be found in Appendix A.

Proximity values between all possible pairs of concepts will be obtained through a paired comparison technique. In the paired comparison technique, participants will rate each pair of concepts on a 1 to 9 scale as to the degree of similarity between the two concepts in that pair. Once all pairs have been rated, distance estimates are computed as the inverse of the similarity ratings. The matrix of distance estimates can then be analyzed to determine the underlying knowledge structure for each participant.

It is advisable to utilize multiple scaling techniques in the knowledge representation stage (Cooke & McDonald, 1987). This study will incorporate multidimensional scaling (MDS) and Pathfinder techniques to analyze the participant's structural knowledge. Although each of these techniques uses the proximity value matrix to analyze the knowledge structures, each technique emphasizes different representations of the data (Cooke & McDonald, 1991; Gonzalvo, Cañas, & Bajo, 1994). Pathfinder captures information about local relationships (i.e., pairs of items that are highly related). MDS captures information about global relations among the set of concepts as a whole (i.e., dimensions). Furthermore, MDS and Pathfinder differ in terms of the type of representation generated (e.g., hierarchical in MDS, network in Pathfinder).

MDS is a powerful technique for extracting the latent structure within the empirical similarity/proximity judgments. This is accomplished by arranging concepts in n-

dimensional space where the distances between points reflect the psychological proximity of the concepts. MDS supplies several useful pieces of information. It summarizes the data into a spatial configuration, captures the global relations among the concepts, and supplies a metric (distance between concepts in multidimensional space) that allows for a quantitative comparison across participant knowledge structures (Schvaneveldt et al., 1985).

Pathfinder produces a network with concepts presented as nodes and relations between concepts are represented as links connecting the nodes. These links may be either directed (traveling in only one direction) or undirected (allowing travel in either direction; Dearholt & Schvaneveldt, 1990). After applying the Pathfinder algorithm, a link remains in the network if and only if that link is a minimum-length path between the two concepts. The length of a path is a function of the weights associated with the links in the path. Different functions for computing path length yield different networks. Some argue (e.g., Schvaneveldt et al., 1985) that Pathfinder is better able to reflect psychological proximity between the concepts because it extracts the latent structure rather than transforming the data, as is done in MDS procedures. On the other hand, Pathfinder is not capable of depicting global relationships, which is the strong point of MDS procedures. For these reasons, both knowledge representation procedures will be used in this study.

Behavior - In order to determine whether the knowledge and skills learned in training has been transferred to the job, behavioral measures will be collected. The knowledge and skill imparted in GUIDE is best assessed by evaluating courseware designed by the participant before and after interacting with GUIDE. A courseware sample will be obtained from each participant before and after interacting with GUIDE. Each courseware sample will

be rated by a panel of research psychologists/scientists as to the extent that the courseware incorporates Gagné's nine events of instruction.

Procedures

An initial courseware sample for each participant will be obtained prior to being exposed to GUIDE. Participants' prior knowledge of instructional design will be determined by administering the paper-and-pencil knowledge test and collecting proximity values between the concepts to assess the novice mental models of instructional design. Proximity values for the list of concepts will be collected from expert instructional designers in order to determine the expert mental model of instructional design with which to compare the mental models of novice instructional designers. After the novice instructional designers have had adequate time (approximately three months) to interact with GUIDE, reactions to GUIDE, as well as a second administration of the two learning measures (paper-and-pencil knowledge test and proximity values between concepts) will be collected. The participants will then have an additional 30 days in which to submit their second courseware sample for evaluation.

References

Adelson, B. (1981). Problem solving and the development of abstract categories in programming languages. *Memory and Cognition*, *9*, 422-433.

Asiu, B. (1995). *GAIDA Focus Group Activity*. Unpublished report. Brooks AFB, TX: Armstrong Laboratory, Technical Training Research Division, Human Resources Directorate.

Bloom, B. (1956). *Taxonomy of educational objectives: The cognitive domain.* New York: Donald McKay.

Carroll, J. M. & Olson, J. R. (1988). Mental models in human-computer interaction. In M. Helander (Ed.) *Handbook of Human-Computer Interaction*. Amsterdam: Elsevier (North-Holland).

Chase, W. G. & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4*, 55-81.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121-152.

Cooke, N. M., & McDonald, J. E. (1987). The application of psychological scaling techniques to knowledge elicitation for knowledge-based systems. International Journal of Man-Machine Studies, 26, 533-500.

Dearholt, D. W. & Schvaneveldt, R. W. (1990). Properties of pathfinder networks. In R. W. Schvaneveldt (Ed.) *Pathfinder Associative Networks: Studies in Knowledge Organization*. Norwood, NJ: Ablex Publishing Corporation.

Dorsey, D. W. & Foster, L. L. (1996). An application of knowledge structure assessment in the computer knowledge domain: Relations with experience and implications for training. Symposium paper presented in R. Klimoski's (chair) *Knowledge Structures and Training: Measurement and Validity Issues*. 11th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, California.

Gagné, R. M. (1984). Learning outcomes and their effects. *American Psychologist, 39*, 377-385.

Gagné, R. M. (1985). *The conditions of learning* (4th ed.). New York, NY: Holt, Rinehart, & Winston.

Gagné, R. M. (1992). *Tryout of an organizing strategy for lesson design: Maintenance procedure with checklist*. Brooks AFB, TX: Armstrong Laboratory, Human Resources Directorate (AL-TP-1992-0016).

Gettman, D. J. (1995). *The GAIDAd approach to instructional design advising: Theory, history, and research*. Unpublished manuscript. Brooks AFB, TX: Technical Training Research Division, Human Resources Directorate.

Glaser, R., & Chi, M. T. (1989). Overview. In M. T. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise* (pp. xv-xxviii). Hillsdale, NJ: Erlbaum.

Goldsmith, T. E., Johnson, P. J., & Acton, W. H. (1991). Assessing structural knowledge. *Journal of Educational Psychology, 83*, 88-96.

Goldstein, I. L. (1991). Training in work organizations. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 2, pp. 507-620). Palo Alto, CA: Consulting Psychologists Press.

Gonzalvo, P., Cañas, J. J., & Bajo, M. (1994). Structural representations in knowledge acquisition. *Journal of Educational Psychology, 85*, 601-616.

Johnson-Laird, P. (1983). *Mental models.* Cambridge, MA: Harvard University Press.

Kirkpatrick, D. L. (1987). Evaluation of training. In R. L. Craig (Ed.), *Training and Development Handbook: A GAIDA to human resource development* (3rd ed., pp. 301-319). New York, NY: McGraw Hill.

Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology, 78*, 311-328.

McKeithen, K. B., Reitman, J. S., Rueter, H. H. & Hirtle, S. C. (1981). Knowledge organization and skill differences in computer programmers. *Cognitive Psychology, 13*, 307-325.

Rouse, W. B., & Morris, N. M. (1986). On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin, 100*, 349-363.

Schvaneveldt, R. W., Durso, F. T., Goldsmith, T. E., Breen, T. J., Cooke, N. M., Tucker, R. G., & De Maio, J. C. (1985). Measuring the structure of expertise. *International Journal of Man-Machine Studies, 23*, 699-728.

Tennyson, R. D. & Gettman, D. J. (1995). *GAIDAd Approach to Instructional Design Advising (GAIDA) evaluation: Student learning, quality, attitudes, and perceptions.* (AL-TP-1995- ??). Brooks AFB, TX: Armstrong Laboratory, Technical Training Research Division, Human Resources Directorate.

Wexley, K. N., & Latham, G. P. (1991). *Developing and training human resources in organizations*. New York, NY: HarperCollins Publishers Inc.

Appendix A

Concepts

GAIDA

Nine Events of Instruction

Computer Based Instruction

Target Audience

Foundation

Implementation

Production

Prior Knowledge

Performance

Transfer

Open-Ended Learning

Cases

Gagné

Instructional Design

Lesson Planning

Learning

Maintenance

Design

Situational Evaluation

Simulation

Retention

Courseware

Examples

STATIC ANTHROPOMETRIC VALIDATION OF DEPTH

Kristie J. Nemeth
Graduate Research Assistant
Center for Ergonomic Research

Department of Psychology
Miami University
Oxford, Ohio 45056

Final Report for:
Summer Graduate Student Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Armstrong Laboratory

September 1996

STATIC ANTHROPOMETRIC VALIDATION OF DEPTH

Kristie J. Nemeth
Center for Ergonomic Research, Miami University, Oxford Ohio

Abstract

The current project examines the validity of the human model in the Design and Evaluation of Personnel, Training and Human Factors (DEPTH) application. A total of 28 anthropometric measurements taken from 17 human volunteers were compared to the measurements taken from the corresponding human model in DEPTH. Although a few dimensions were accurately represented, many had large discrepancies. Some of the measurement deviations could be explained by different measurement procedures, but this cannot account for all of the error. Given the large discrepancy found in the hand and forearm sections of the body, the current version of DEPTH would not allow a designer to accurately simulate reach and grasping tasks. Future research should continue to consider additional anthropometric dimensions which are necessary to realistically simulate a human figure. In addition to static body measurements, it is necessary to examine the human body in motion. To accurately simulate a human performing a task, information about the size, shape and movement of the model are necessary.

STATIC ANTHROPOMETRIC VALIDATION OF DEPTH

Kristie J. Nemeth
Center for Ergonomic Research, Miami University, Oxford Ohio

Introduction

The goals for the Design and Evaluation of Personnel, Training and Human Factors (DEPTH) application include the ability to illustrate, predict, evaluate and describe human-machine interactions in a simulated environment. It is expected that a designer could accurately measure a potential user, and create a human-computer-model (henceforth known as model) based on this individual. If the model is able to reach, lift and fit in the virtual workspace, than the human should also be able to perform the same tasks in the real workspace. The flexibility of computer simulations to perform engineering and ergonomic analysis saves time and money because designers can spot potential problems prior to physical mock-up.

Unfortunately, the human body and its' behavior is a complicated system. It is not a simple task to create a simulated human-model which looks like and acts like a real human being. Many issues need to be addressed, such as size, movement, mass and strength, and cognitive components (like intention). Before the DEPTH software is widely used in the development of work environments, a validation of the model sizing is required.

The first concern would be to verify that the measurements taken from a human are accurately reflected in the resulting model. Unless the size of the model approximates the intended human, any workplace evaluations based on the model must be suspect. Of course, some variations are to be expected because human bodies are not made of simple linkages. The hip and shoulder are examples of complicated joints which can only be approximated.

In addition to single segments, the combination of body segments should be investigated. By placing the human and model into standard postures we can compare specific measurements. For example, stature is a dimension which includes many body segments - upper and lower legs, trunk, neck and head. The stature of the model should be accurate if these segments are combined correctly.

The current project examines the validity of the DEPTH model. A total of 65 anthropometric measurements were taken from 17 human volunteers. Each dimension was chosen with the goal of quantifying the width, length and depth of each body segment using the existing measurement equipment. The list of these dimensions can be found in Appendix A. The measurement procedure for 28 of these dimensions corresponded to the written description of the automatic measurements available in DEPTH (automatic measures defined in Appendix B). Analyses were then performed on the differences between the human's and model's measurements for each of the twenty-eight dimensions.

Method

Participants: Volunteers from AL/HRG were solicited. Each volunteer was asked to wear clothing (such as shorts and t-shirt) which allowed access to body landmarks. A private room was made available to allow the participant to change clothing and be measured by the experimenter. A female experimenter carried out measurements on both male and female participants.

Materials: A measuring device was used which included the functions of an anthropometer, beam and spreading caliper. A tape measure was used to measure circumference, and a wall scale for reach dimensions.

Procedure: After describing the purpose of the project and obtaining an informed consent, each participant was asked general information questions. Then they assumed a variety of postures (i.e. sitting, standing) and measurements were taken for each of the dimensions listed in Appendix A. The standard protocol for each dimension was used (Gordon, et al, 1989).

Once the body measurements were completed, a computer-model was developed for each participant using DEPTH v4.1 (similar results were obtained using DEPTH v4.2). The automatic measurement procedure was then used to quantify the anthropometric dimensions of the model.

Results

An analysis was conducted to compare the actual human's measurements to the model's measurements across all dimensions. A positive correlation, r=.9849, demonstrates a strong relationship between the human and corresponding model's measurements. In Figure 1, the solid line represents a perfect correlation (y=x). The broken line demonstrates the actual relationship (y=0.95+2.17). Although there is a strong correlation between the measures, the smaller dimensions tend to be overestimated, while larger dimensions are underestimated.

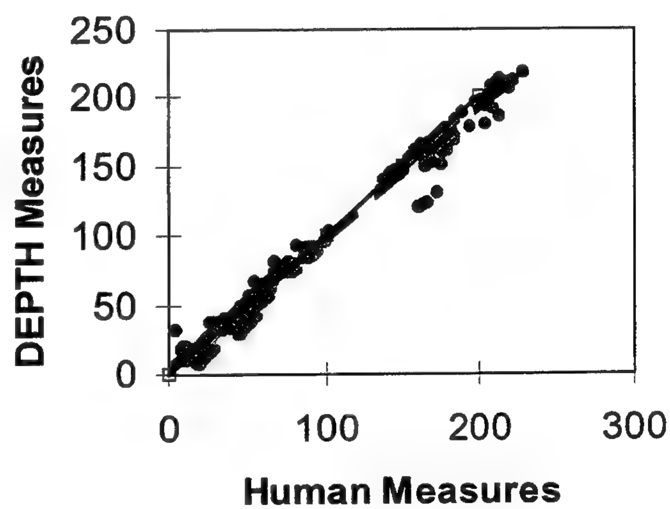


Figure 1. The solid line represents a perfect correlation, while the broken line shows the actual correlation between Human and DEPTH measures.

Because this analysis did not fully characterize our observations of how well the model reflected the human, additional analyses were conducted to further examine the relationship. The pattern of results for single-segment and segment-combined dimensions are comparable, and therefore combined in the following analyses.

Table 1. **The estimation and percent error is shown for each dimension.**

Dimensions	Human	DEPTH	Est. error	%Error
Acromial Height Sitting	60.15	63.31	3.16	5.2%
Acromial Height Standing	143.68	144.68	1.00	0.7%
Biacromial Breadth	35.69	39.16	3.47	9.7%
Biceps Circumference	32.49	37.77	5.28	16.2%
Bideltoid Breadth	47.51	46.68	-0.83	1.8%
Buttock-Knee Length	58.89	60.7	1.81	3.1%
Buttock-Popliteal Length	48.25	48.75	0.50	1.0%
Calf Circumference	37.56	37.85	0.29	0.8%
Chest Depth	23.44	25.36	1.92	8.2%
Eye Height Standing	159.98	159.19	-0.79	0.5%
Eye Height Sitting	77.38	78.42	1.04	1.3%
Foot Breadth	9.79	10.46	0.67	6.8%
Foot Length	26.38	28.26	1.88	7.1%
Head Circumference	58.06	54.62	-3.44	5.9%
Hip Breadth Sitting	38.99	35.41	-3.58	9.2%
Hip Breadth Standing	35.27	35.41	0.14	0.4%
Knee Height Sitting	53.78	55.55	1.77	3.3%
Neck Circumference	37.31	37.79	0.48	1.3%
Popliteal Height	43.08	40.84	-2.24	5.2%
Shoulder-Elbow Length	34.37	34.72	0.35	1.0%
Stature Sitting	90.79	89.93	-0.86	0.9%
Stature Standing	173.39	173.04	-0.35	0.2%
Thigh Circumference	58.27	55.96	-2.31	4.0%
Thigh Clearance	15.04	18.84	3.80	25.3%
Dimensions related to the hand/forearm region:				
Forearm-Hand Length	43.24	31	-12.24	28.3%
Hand Breadth	8.27	18.82	10.55	127.6%
Hand Length	18.88	9.19	-9.69	51.3%
Overhead Reach	213.32	208.16	-5.16	2.4%
Span	175.30	156.61	-18.69	10.7%

Note: all measurements in centimeters

Initial observations of the computer models showed gross errors in the representation of the forearm and hand sections. For this reason, the 5 dimensions related to these two body sections are handled separately. An alpha level of .05 was used throughout this study. Group comparisons were done using the Tukey HSD test. All cell means are shown in Table 1.

The raw anthropometric measurements were subjected to a two-way repeated measures analysis of variance with two levels of source (human and computer model), and 26 dimensions (see Appendix A). A significant interaction was found between source and dimension, $\underline{F}(25, 399)=8.76$, $\underline{p}=.0001$. A significant Main

Effect of dimension demonstrates only that some dimensions measure large areas while others measure only small body parts, $F(25, 400)=3781.78$, $p=.0001$. The Main Effect of Source approaches significant levels to show that the model's measurements tend to overestimate the human's measurements, $F(1, 16)=3.82$, $p=.0683$.

The raw data for the 5 hand/forearm-related dimensions were used in a similar analysis. A significant interaction was found between source and dimension, $F(4, 94)=99999.99$, $p=.0001$. A significant Main Effect of dimension also shows only that some dimensions measure large areas while others measure only small body parts, $F(4, 96)=1982.83$, $p=.0001$. The Main Effect of Source shows that the model's measurements significantly underestimate the human's measurements, $F(1, 16)=6.58$, $p=.0208$.

Estimation errors (difference between the human and model's measurements) were subjected to a one-way repeated measures analysis of variance with the 26 dimensions not related to hand/forearm sections. A significant difference was found between the estimations error for each dimension, $F(25, 399)=8.71$, $p=.0001$. The same effect was found for the 5 hand/forearm-related dimensions, $F(4, 399)=50.49$, $p=.0001$. These results show that the magnitude of the measurement difference are not consistent. Some dimensions are underestimated, and others overestimated.

Some of the anthropometric dimensions are clearly larger than others, and this may confound the issue of comparing estimation errors. It is more important if the knee height is off by 5 cm than if stature was off by the same amount. To compensate for this effect, the next analysis considers the percent error. The estimation error was compared to the human's actual measurement for each dimension to compute the percent error.

A one-way repeated measures analysis of variance with the 26 dimensions not related to hand/forearm sections found that there were significant differences in percent error between dimensions, $F(25, 399)=31.09$, $p=.0001$. The overlapping groupings of dimensions that show significantly different percent errors, does not clearly show a relevant cutoff point. For example, the percent error for Foot Breadth (6.8%) is significantly greater than that for Bideltoid Breadth (-1.4%), while Knee Height (3.3%), seated is significantly greater than that of Thigh Circumference (-3.69). It is not clear where to draw a line to determine "small" and "large" percent estimation error. The range of percent error in each of the given cases is greater than 5%, but it is not clear whether "acceptable error" could be in the range of +/-2.5% (5% total error acceptable) or +/-5% (10% total error acceptable).

A significant difference was also found for the 5 hand/forearm-related dimensions, but this may be due, in part, to the extraordinarily large error in hand breadth, , $F(4, 399)=1255.49$, $p=.0001$. For these dimensions also it is not clear how "acceptable error" should be defined. This issue is further discussed in the next section.

Discussion

The problem now is how to define a "realistic" computer model. Although several workplace-simulation applications exists, there is little documentation on this issue. For many of the dimensions examined in this project, statistically significant differences were found between the human's and the model's measurements. Is this

the appropriate means to finding a simulation acceptable, or unacceptable? Another method would be to instead consider practical differences, except that the acceptable percent-error cut-off point would be arbitrary. If a suggested 5% error is considered acceptable, the model can either under or overestimate human size by 5%, for a total of 10% "acceptable error". While 5% error would certainly be satisfactory for hand length (off by +/- 1cm), the same percent error for acromion height would be more significant (off by +/- 7cm). When an aircraft maintenance worker is trying to repair and replace crowded components, a few centimeters may make the difference between using the correct tool and finding the component inaccessible.

Due to this uncertainty in validation procedures for human-machine simulation software, the final word on the validation of the static anthropometry of the DEPTH models is not clear. Although the stature of the model accurately represents the stature of the corresponding human, with less than ½% error, biacromial breadth (shoulder width) is off by nearly 10%. The hand of the computer model is on average over twice the width and half the length of the human's actual hand. The forearm on the computer models ranges from 21% to 36% underestimation (average difference of error is 12.24cm). Given the large discrepancy found in the hand and forearm sections of the body, the current version of DEPTH would not allow a designer to accurately simulate reach and grasping tasks. Any design test which involved reaching for a component, or accessing an element of the work area would be significantly inaccurate.

The list of dimensions used in this project is not intended to be a comprehensive analysis of the DEPTH model. The dimensions were chosen because they were both relevant to the anthropometric characterization of a model and feasible to evaluate with the current software and measurement equipment. Also, it is not currently possible to verify the definitions of the automatic DEPTH measurements. Although the descriptions are similar, it is possible that some of the error can be explained by different measurement procedures.

Another thing to consider is that a few dimensions with large percent errors may not be as relevant to maintenance tasks (i.e., thigh clearance = 25% overestimation). However, they should be considered if the DEPTH software is to ever be used in other applications (i.e., cockpit, or office environments). Future research should continue to consider additional anthropometric dimensions which are necessary to realistically simulate a human figure. In addition to static body measurements, it is necessary to examine the human body in motion. To accurately simulate a human performing a task, information about the size, shape and movement of the model are necessary.

APPENDIX A

Measures used in analysis:

Unrelated to hand/forearm

Acromial Height Sitting
Acromial Height Standing
Biacromial Breadth
Biceps Circumference Flexed
Bideltoid Breadth
Buttock Knee Length
Buttock Popliteal Length
Calf Circumference
Chest Depth
Eye Height Sitting
Eye Height Standing
Foot Breadth Horizontal
Foot Length
Head Circumference
Hip Breadth
Hip Breadth Sitting
Knee Height Sitting
Neck Circumference
Popliteal Height
Stature Sitting
Stature Standing
Thigh Circumference
Thigh Clearance from table

Related to the hand/forearm

Hand Breadth
Hand Length
Overhead Fingertip Reach
Shoulder Elbow Length
Span

Additional measurements not in analysis:

Abdominal Ext Depth Sit
Acromion Radiale Length
Ball Of Foot Length
Bitragion Breadth
Bitragion Coronal Arc
Buttock Depth
Cervicale Height
Cervicale Height Sitting
Chest Breadth
Elbow Rest Height
Forearm Circumference Flexed
Head Breadth
Head Length
Iliocristale Height
Knee Height Midpatella
Lateral Femoral Epicond Ht
Lateral Malleolus Height
Lower Thigh Circumference
Neck Circumference (Base)
Overhead Fingertip Reach Ext
Overhead Fingertip Reach Sit
Radiale Stylion Length
Sleeve Length Outseam
Tenth Rib Height
Thigh Clearance from floor
Thumb Tip Reach
Tragion Top of Head
Trochanteric Height
Waist Breadth
Waist Depth
Wrist Center Of Grip Length
Wrist Height
Wrist Height Sitting
Wrist Index Finger Length
Wrist Thumb Tip Length
Wrist Wall Length
Wrist Wall Length Extend

APPENDIX B

AutoDEPTH Dimensions	Definition for AutoDEPTH measurement
Anterior Arm	arm straight in front of body, measured from back of shoulder to tip of thumb
Arm Span	arms extended out of shoulder, distance between tips of middle fingers
Biceps	maximum circumference
Bideltoid	maximum breadth across the upper arms
Buttock-Knee	sitting, distance from rearmost projection to the front of right kneecap
Buttock-Popliteal	sitting, distance from rearmost projection to the back of right kneecap
Calf	maximum circumference of right calf
Chest Depth	maximum thickness of chest
Eye Sitting	distance from sitting surface to inner core of right eye
Eye Standing	distance from floor to inner core of right eye
Foot Breadth	maximum breadth of right foot
Foot Length	distance from back of right heel to tip of longest toe
Forearm-Hand	distance from elbow to tip of middle finger
Functional	distance from back of shoulder to tip of thumb
Hand Breadth	maximum breadth across base of fingers
Hand Length	from wrist to tip of middle finger
Head Circumference.	Maximum circumference of head
Head Length	menton to top of head
Hip Sitting	sitting, maximum breadth across the hips
Hip Standing	standing, maximum breadth across the hips
Knee Sitting	sitting, from floor to top of knee cap
Neck	maximum circumference of neck
Overhead	arm extended above shoulder, distance from floor to tip of middle finger
Popliteal	distance from surface of footrest to underside of right knee
Shoulder Breadth	biacromial breadth - maximum breadth across shoulders
Shoulder Sitting	sitting, distance from seated surface to right shoulder
Shoulder Standing	standing, distance from floor to outer point of shoulder
Shoulder-Elbow	distance from outer point of shoulder to elbow
Stature Sitting	sitting, distance from seated surface to top of head
Stature Standing	standing, distance from floor to top of head
Thigh	maximum circumference of thigh
Thigh Clearance	distance from top of sitting surface to junction abdomen-thigh

EVALUATION OF VARIOUS SOLVENTS FOR THE USE IN A NEW SAMPLING DEVICE FOR THE COLLECTION OF ISOCYANATES DURING SPRAY-PAINTING OPERATIONS

Samuel H. Norman
Graduate Student
Department of Chemistry

Southwest Texas State University
San Marcos, Texas 78666

Final Report for:
Graduate Student Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Armstrong Laboratory

August 1996

EVALUATION OF VARIOUS SOLVENTS FOR THE USE IN A NEW SAMPLING DEVICE FOR THE COLLECTION OF ISOCYANATES DURING SPRAY-PAINTING OPERATIONS

Samuel H. Norman
Graduate Student
Department of Chemistry
Southwest Texas State University

Abstract

The properties of several solvents were evaluated for their possible use as a collection media, together with derivatizing reagent, for polyisocyanates in a newly developed sponge type air sampler. Several solvent-sponge properties were tested which include sponge expansion after soaked in the solvent, extraction from the sponge by the solvent, loss of any solvent from the sponge during air flow, and polyisocyante recovery from the sponge after sampling. The solvents tested were acetonitrile, toluene, tributylphosphate, butyl benzoate, mesitylene, acetophenone, benzyl ether, benzyl ethyl ether, 2-nitro-m-xylene, and phenetole. Acetonitrile and toluene were very good solvents for preparing polyisocyanate standards, however due to there high volatility could not be used. Acetonitrile was chosen as the solvent of choice for extraction of derivatized polyisocyanate after sampling. Tributylphosphate interfered with the reaction between derivatizing reagent and polyisocyanate and could not be used. Mesitylene, benzyl ether, and benzyl ethyl ether extracted interferants from the sponge thereby rendering each one useless as a working solvent. Butyl benzoate and phenetoleare moderately volatile and could have been considered, however no polyisocyanate recovery from the sponge was detected. Acetophenone and 2-nitr-m-xylene exhibited all of the desirable properties. These two solvents were usable and were chosen for future study as working solvents in the new sponge sampling device.

EVALUATION OF VARIOUS SOLVENTS FOR THE USE IN A NEW SAMPLING DEVICE FOR THE COLLECTION OF ISOCYANATES DURING SPRAY-PAINTING OPERATIONS

Samuel H. Norman

Introduction

Isocyanates are major components in polyurethane paint formulations because of their ability to form durable crosslinks in the polyurethane coating. The most popular isocyanate employed for this hardening is 1,6-hexamethylene diisocyanate (HDI) in its oligomeric form. The monomeric form of HDI was used in the past, but due to its toxicity, it has been highly regulated by the American Council of Governmental Industrial Hygienists (ACGIH).[1] Because regulations were imposed on the use of HDI monomer as a paint hardener, industries started using the HDI oligomer (mainly biuret or isocyanurate of HDI) as the primary hardener in paint formulations. The oligomers were thought to be less harmful because they were less volatile than the monomer and thus less accessible during respiration. However, recent reports have indicated that HDI oligomer, or polyisocyanates, can cause occupational asthma as well as other respiratory problems for workers involved in spray-painting operations.[2-4]

Because of these health risks involved with the use of polyisocyanates in spray paint formulations, routine chemical analysis of the concentration level of airborne polyisocyanates during spray-painting operations will be needed. An important factor in isocyanate analysis is the complete and efficient collection of isocyanate from the atmosphere where it may be present. This factor involves the use of a device that can capture all isocyanates and will give a proper representation of the isocyanate air concentration. The first sampling devices that were designed were for the isocyanate monomers. Because most of the monomers exist in the vapor phase after they are sprayed, a device that will collect vapors efficiently is needed. In industrial hygiene monitoring, generally, impinger collection techniques are recommended for sampling contaminants that are normally present as vapors.[5] Contaminated workplace air is pulled through an impinger that contains a derivatizing reagent dissolved in an absorbing solvent. The isocyanate vapors present in the air absorb into the solvent and are derivatized by the reagent. Airborne polyisocyanates, however, have very low vapor pressures at normal ambient temperatures and would be present as condensation aerosol (general size range 0.01 to 1 μm).[5] Impinger techniques have been shown to give poor collection efficiency for particles less than about 2 μm in size.[5]

Therefore, a device which is capable of efficiently collecting aerosol particles as well as monomer vapors is needed. Techniques which involve the use of a glass fiber filter that is impregnated with derivatizing reagent have been used to collect aerosol particles, however isocyanates are extremely reactive and react with other compounds in the filter or other particles that have been collected. A possible technique that may satisfactorily collect all airborne isocyanate which involves the combination of an impinger followed by a reagent-coated filter has been proposed.(6) However, this poses difficulties in setting up equipment and personal sampling. A novel approach to sampling both isocyanate vapors and aerosols involves the collection of contaminated air using a cartridge sampler. The cartridge is a small cylinder through which air can be drawn and contains a portion of a polyester type polyurethane foam that is saturated with solvent and derivatizing reagent. The backing of the cartridge is a reagent impregnated glass fiber fiter followed by an alumunum mesh screen for support. The resulting sampling device seems capable of trapping and derivatizing aerosol particles at the polymer backing while vapors are collected and derivatized in the saturated polyurethane foam.

In the following report, certain characteristics of a polyurethane foam were tested to construct a sampling device for the collection of airborne polyisocyanates. The following foam characteristics were studied: solvent-foam combatibility, loss of solvent from foam during air sampling, solvent-foam extractions, recovery of polyisocyanate urea from foam using the different solvents, and foam back-up efficiency.

Experimental

Materials

Polyurethane sponges were obtained from Polyplastics (10201 Metropolitan Dr., Austin, TX). The polyester type polyurethane (PUF) sponge (4#CE) chosen for further evaluation was in the form of a 2" thick, 12" x 12" block. The 25 mm glass fiber filters (GFF) were obtained from Omega Specialty Instrument Co. The aluminum mesh screen (1 mm^2 openings) was obtained at Ace Hardware and cut to fit. The cartridge cassette samplers chosen were asbestos samplers obtained from Nucleopore (Pleasanton, CA).

Sponge Evaluation

Open-celled, porous PUF sponges were cut into cylinders (2.5 cm in diameter and 2.5 cm in height) using a sharp-edged carbon-tipped die (hole saw, 1 1/8" Fort Worth, IN) and scissors. These sponges were soaked overnight in acetonitrile (HPLC grade, EM Science) to remove

impurities and dried in an oven at 110 C. However, due to later results, the sponges were allowed to air-dry rather than in the oven because of sponge degradation from heating. The sponges were then soaked overnight in toluene, tributylphosphate, butyl benzoate, mesetylene, acetophenone, benzyl ether, benzyl ethyl ether, 2-nitro-m-xylene, and phenetole. All of these solvents were at least reagent grade. Only those sponges which maintained their mechanical rigidity and did not react with solvent were deemed suitable for further study.

Sampler Preparation

The PUF sponge was cut into cylinders (2.5 cm in diameter x 5.0 cm in height) using the sharp-edged carbon-tipped die. The cylinders were then cut to a height of 2.5 cm, soaked in a solution of derivatizing reagent in solvent (~500 mg/L) and squeezed repeatedly until just moist. The glass fiber filters were soaked in a solution of derivatizing reagent in acetonitrile (~500 mg/L) and allowed to air-dry. The aluminum mesh screen was cut to a diameter of 2.5 cm with the same die.

The cellulose acetate filter was removed from the lower portion of the Nucleopore asbestos sampler and replaced by the aluminum mesh screen. The prepared glass fiber filter was placed on top of the mesh screen, and the lower portion of the cartridge attached tightly to the body of the cassette sampler. The moist PUF sponge was then inserted into the cassette body on top of the glass fiber filter and the upper portion of the sampler was tightly attached to the body. The supplied stopper pins (red and blue) were positioned into the inlet and outlet of the cassette sampler.

Reagents and Apparatus

The derivatizing reagent, 1-(2-methoxyphenyl)piperazine (MOP), was obtained from Fluka. Desmodur N-75, which is 75% HDI polyisocyanate in xylene and contains 35-40% biuret of HDI (MSDS), was obtained from Miles Inc. (Pittsburgh, PA). All other chemicals and solvents were reagent grade or better.

The HPLC system consisted of a Hewlett-Packard Series 1090 chromatograph with autosampler and diode array UV-VIS detector operated at 246 nm for MOP. A Hewlett-Packard 1049A electrochemical detector operated in the amperometric mode at +0.8 V was connected in series with the HPLC when possible. Hewlett-Packard 3396 series II integrators were used to determine the area under all chromatographic peaks.

A Supelco Supelcosil LC-8-DB, 3 µm (75 X 4.6 mm) column and various acetonitrile/methanolic buffer (pH=6.0) mobile phases were used for analyzing the MOP

derivatives. A Hewlett-Packard LiChrosorb RP-18, 10 μm (200 X 4.6 mm) column and various acetonitrile/methanolic buffer (pH=6.0) mobile phases were also used for analyzing the MOP derivatives. The mobile phase flow rate for the HDI polyisocyanate derivatives was 1.000 mL/min. The sample injection volume was consistently 20 μL.Dupont Alpha 1 pumps were used to draw air through the sampling devices. The pumps were calibrated before and after air sampling with a Gillian bubble generator.

A Waters Quanta 4000 capillary zone electrophoresis system equiped with a Waters 730 data module was used to analyze for MOP-polyisocyanate urea when using solvents that abrb in the UV region of interest (242 nm). All samples were injected using hydrostatic injection for 10 s. The capillary column had an effective length of 49.0 cm and an inner diameter of 75 μm. The total column length was 46.5 cm. The operating voltage was set at 20 kV and a 0.015 M sodium phosphate buffer adjusted to a pH of 3.0 was used. The detector wavelength was set at 254 nm. The column was reconditioned daily with 0.5 M KOH, and the capillary was purged with buffer for 2 min before each analysis.

Description of Recovery Studies

A standard ~10,000 μg/mL stock solution of the polyisocyanate was prepared from Desmodur N-75 by dissolving the polyisocyanate in toluene. The derivatizing reagent solution was made by dissolving ~50 mg of MOP in 100 mL of the solvent which was being evaluated.

The polyester-polyurethane sponge (#4 CE) was soaked in the derivatizing reagent solution, then squeezed until just moist (approximately 0.8 - 2.5 mL of derivatizing solution remained in the sponge depending on the solvent being used). The sponge was positioned in either a 30 mL beaker to test for recovery from the sponge or within the sampler after assembly of the cassette to test the sampling method.

Variable amounts of polyisocyanate solution were added to the sponge depending on the desired concentration (usually 20 ppm). Air was drawn for 30 minutes through the sponges that were positioned within the cassette samplers while the sponges in the beakers were allowed to sit for 30 minutes. Sponges were then soaked with about 23 mL of a 500 mg/mL solution of MOP in acetonitrile, massaged repeatedly with a stirring rod, and an aliquot then removed from the saturated sponge. 25 - 60 μL of acetic anhydride was added, the solution filtered through a 0.45 μ Nylon filter to remove any particulates, then run on the HPLC.

Polyisocyanate standards were prepared in the same manner as the sponge recovery studies, but without the sponge present in the solution.

Analysis

The results obtained for HDI-based polyisocyanate using the MOP derivatizing reagent were quantitated using the UV detector response when non-absorbing solvents were used. When absorbing solvents were used in the analysis, the results obtained for HDI-based polyisocyanate using MOP were quantitated using the electrochemical detector (ECD) response or the CZE UV detector response. The ECD response was also used to confirm the presence of isocyanates or moieties that contain isocyanate groups when the UV detector response was used.

The recovery and concentration of polyisocyanate in the recovery studies was determined by using the area under the peak in the chromatogram of a recovery sample and comparing it with the area under the corresponding peak in the chromatogram of a standard. Any calibration curves based on MOP-polyisocyanate urea standards were all linear with a correlation coefficient of $r^2 > 0.9975$.

Results and Discussion

Table I on page 23-8 represents some of the general characteristics that were observed between a certain solvent and the foam sponge. Sponge expansion describes the degree of solvent absorption by the sponge where minimal denotes no observed expansion of the sponge, moderate denotes expansion of ~1 cm greater than the original diameter, and strong denotes an expansion of greater than ~1 cm past the original diameter. Extraction from the sponge describes any detectable amount of interferants extracted from the sponge by the solvent. Loss of solvent during sampling describes the percent loss of solvent after air was flowed through the sponge/cassette at 1 L/min for 30 min. Polyisocyanate recovery describes the quality of polyisocyanate recovery from the sponge after a sampling test.

TABLE I. Observed Solvent-Sponge Characteristics

Solvent	Sponge Expansion	Extraction from Sponge	Loss of Solvent During Sampling	Polyisocyanate Recovery
Acetonitrile	Moderate	n.d.	98 %	n.d.
Toluene	Minimal	n.d.	96 %	n.d.
Tributylphosphate	Minimal	n.d.	8 %	n.d.
Butyl Benzoate	Minimal	n.d.	----	n.d.
Mesitylene	Minimal	Strong	36 %	n.d.
Acetophenone	Strong	n.d.	9 %	good
Benzyl Ether	Moderate	Moderate	2 %	----
Benzyl Ethyl Ether	Moderate	Moderate	16 %	----
2-Nitro-m-Xylene	Minimal	n.d.	6%	good
Phenetole	Moderate	Minimal	26 %	n.d.

n.d. = none detected

Acetonitrile

Considerable expansion of the sponge was observed when it was soaked with acetonitrile, however the sponge seemed quite stable. Standards were easily prepared using this solvent, however no detectable polyisocyanate recovery was observed which is probably due to the high volatility of acetonitrile; almost all of the acetonitrile evaporated during sampling (air flow). Acetonitrile extracted nothing from the sponge and was chosen to be used as the solvent for initial cleaning of the sponges prior to sampling; the high volatility of the solvent allows one to air-dry the sponges. Furthermore, because of the above properties and the fact that all of the other solvents are very soluble in it, acetonitrile was chosen to be the primary extracting solvent used to extract the derivatized polyisocyanate in the sponge.

Toluene

Very little expansion of the sponge was observed when it was soaked in toluene and the stability of the sponge was suitable. Standards were easily prepared using toluene, however, no detectable polyisocyanate recovery was observed which is probably due to toluene's high volatility;

almost all of the toluene evaporated during sampling (air flow). There was no detectable extraction of interferants from the sponge and could be used as an extracting solvent to extract the derivatized polyisocyanate from the sponge. Acetonitrile was chosen as the extracting solvent over toluene because acetonitrile does not absorb on the UV detector, whereas toluene does absorb and could be present as an interfering species.

Tributylphosphate

Very little expansion was observed when the sponge was soaked in tributylphosphate and the stability of the sponge was good. The usual properties observed with each of the solvents rated this solvent very high; low volatility, no detectable extraction from sponge, minimal loss of solvent during air flow, and no UV absorption. Although tributylphosphate seemed to be a good solvent for use in the sampling device, standards could not be prepared using this solvent. When different amounts of tribultylphosphate were added to the MOP-polyisocyanate reaction (0.0 - 4.0 ppm), a trend was observed after analysis. The trend showed that for increasing amounts of solvent, the MOP-polyisocyanate peak decreased. This shows that the solvent interferes with the reaction between the derivatizing reagent and the polyisocyanate and could therefore not be used.

Butyl Benzoate

Minimal expansion was observed when sponge material was soaked in butyl benzoate and it remained intact after it was soaked. Preparation of standards was not attempted due to the lack of interest in the solvent because of its strong UV absorption; this solvent was evaluated before the idea of peak detection by electrochemical activity was decided. Loss of solvent during air flow was not tested, as well as polyisocyanate recovery from the sponge. Low volatility, however, was a good feature for butyl benzoate.

Mesitylene

Very little sponge expansion was observed when the sponge was soked in mesitylene. The stability of the sponge after it was soaked was good, and the moderate volatility of this solvent was a slight disadvantage (36 % loss after air flow). Extraction of interferants from the sponge by the solvent was very strong which rendered mesitylene useless as a solvent for the sampling device. No further testing was done with this solvent.

Acetophenone

There was strong expansion of the sponge when it was soaked in acetophenone, however

the sponge did remain very stable. Because this solvent absorbs strongly in the UV region of interest, electrochemical detection was adopted. There were no detectable extraction interferants using acetophenone, and due to its low volatility only 9 % of the solvent was lost after air was drawn through the sponge. There was polyisocyanate recovery from the sponge after sampling in the assembled sampling device, however limited data on this recovery can only rate it as good recovery. Further studies will be attempted to assess acetophenone as a working solvent for the new sampler. Capillary zone electrophoresis (CZE) with UV detection was also useful to separate and detect the MOP-polyisocyanate urea (CZE parameters are in the experimental section).

Benzyl Ether

Moderate expansion of the sponge was observed after it was soaked in benzyl ether, and the sponge's stability was suitable. This solvent also absorbs in the UV region of interest and electrochemical detection was used. Several interferants that were extracted from the sponge by the solvent were detected which rendered benzyl ether useless. While the other properties of volatility and solvent loss were good, no further tests were done for benzyl ether.

Benzyl Ethyl Ether

Moderate expansion of the sponge was observed after it was soaked in benzyl ethyl ether, and the sponge's stability was suitable. This solvent also absorbs in the UV region of interest and electrochemical detection was used. Several interferants that were extracted from the sponge by the solvent were detected which rendered benzyl ethyl ether useless. While the other properties of volatility and solvent loss were good, no further tests were done for benzyl ethyl ether.

2-Nitro-m-Xylene

There was minimal expansion of the sponge when it was soaked in 2-nitro-m-xylene, and the sponge remained very stable. Because this solvent also absorbs strongly in the UV region of interest, electrochemical detection was used. There were no detectable extraction interferants using 2-nitro-m-xylene, and due to its low volatility only 6 % of the solvent was lost after air was drawn through the sponge. There was polyisocyanate recovery from the sponge after sampling in the assembled sampling device, however limited data on this recovery can only rate it as good recovery. Further studies will be attempted to assess 2-nitro-m-xylene as a working solvent for the new sampler. As with acetophenone, capillary zone electrophoresis with UV detection was useful to separate and detect the MOP-polyisocyanate urea.

Phenetole

There was moderate expansion observed when the sponge material was placed in phenetole, however its stability remained fair. This solvent also absorbs in the UV region of interest and electrochemical detection was used. Some detection of extaction interferants from the sponge were observed, but further tests were completed. Phenetole's moderate volatility allowed a 26 % loss of solvent after drawing air. MOP-polyisocyanate standards were prepared in phenetole, however no detectable recovery from the sponge was observed.

Conclusions

After thourough evaluation of each of the solvents included in this report, acetophenone and 2-nitro-m-xylene were chosen to be the most suitable solvents to be employed in the new sponge air sampler. The two main properties which facilitated this decision were volatility of the solvent and detection of extracted interferants from the sponge by the solvent. While many of the solvents may have allowed good reaction of the derivatizing reagent and polyisocyanate, they were eliminated because of either solvent loss after air flow (high volatility) or the extraction of substances (interferants) from the sponge. The two solvents chosen for further study exhibited the desired properties of low volatility, no detectable extraction interferants, and most importantly polyisocyanate recovery. Another solvent chosen for further study because of its possible properties was dimethyl sulfoxide.

References

1. American Council of Governmental Industrial Hygienists (ACGIH): *Threshhold Limit Values and Biological Exposure Indices for 1986-1987*. Cincinatti, Ohio: ACGIH, 1987.

2. Belin L., U. Hjortsberg, U. Wass: Life-Threatening Pulmonary Reaction to Car Paint Containing a Prepolymerized Isocyanate. Scand. J. Work Environ. Health 7: 310-312 (1981).

3. Malo J. L., G. Ouimet, A. Cartier, D. Levitz, R. Zeiss: Combined Alveolitis and Asthma Due to Hexamethylene Diisocyanate (HDI) with Demonstration of Crossed Respiratory and Immunologic Reactivities to Diphenylmethane Diisocyanate (MDI). J. Allergy Clin. Immunol. 72: 413-419 (1983).

4. Nielsen J., C. Sango, G. Winroth, T. Halberg, S. Skerfving: Systemic Reactions Associated with Polyisocyanate Exposure. Scand. J. Work Environ. Health 11: 51-54 (1985).

5. Dharmarajan, V., R.D. Lingg, K.S. Booth, and D.R. Hackathorn: Recent Developments in the Sampling and Analysis of Isocyanates in Air. *In Sampling and Calibration for Atmospheric Measurements*. J.K. Taylor, ed. [ASTM STP 957] Philadelphia: American Society for Testing and Materials, 1988. pp. 190-202.

6. Adkins, J.E., and N.W. Henry: Industrial Hygiene. *Anal. Chem.* 67: 349R-376R (1995).